

Modelos animales para caracteres con efectos maternos, madres no identificadas e información incompleta.

*Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires,
Área Ciencias Agropecuarias*

María José Suárez
Ingeniera Agrónoma, Facultad de Agronomía, Universidad de Buenos Aires
– 2006 –

Lugar de trabajo: Cátedra de Mejoramiento Genético Animal, Departamento de
Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires



Escuela para Graduados Ing. Agr. Alberto Soriano
Facultad de Agronomía – Universidad de Buenos Aires



COMITÉ CONSEJERO

Director de tesis:

Rodolfo Juan Carlos Cantet

Ingeniero Agrónomo, Universidad de Buenos Aires, Argentina.

MSc., Montana State University, Bozeman, MT., EE.UU.

MSc., University of Illinois, Urbana-Champaign, IL., EE.UU.

Ph.D., University of Illinois, Urbana-Champaign, IL., EE.UU.

Consejera de Estudios:

Mónica Graciela Balzarini

Ingeniera Agrónoma, Universidad Nacional de Córdoba, Argentina.

MSc., Universidad de Buenos Aires, Argentina

Ph.D., Louisiana State University, EE.UU

JURADO DE TESIS

Director de tesis:

Rodolfo Juan Carlos Cantet

Ingeniero Agrónomo, Universidad de Buenos Aires, Argentina.

MSc., Montana State University, Bozeman, MT., EE.UU.

MSc., University of Illinois, Urbana-Champaign, IL., EE.UU.

Ph.D., University of Illinois, Urbana-Champaign, IL., EE.UU.

JURADO

Daniel Omar Maizon

Médico Veterinario, Universidad de Buenos Aires, Argentina.

MSc., Universidad de Buenos Aires, Argentina.

Ph.D., Cornell University, Ithaca, NY., EE.UU.

JURADO

Jorge Ignacio Urioste Aguerre

Ingeniero Agrónomo, Sveriges lantbruksuniversitet, Suecia.

Msc., Sveriges lantbruksuniversitet, Suecia.

Dr., Universidad de la República, Uruguay.

JURADO

Luis Varona Aguado

Licenciado en Veterinaria. Universidad de Zaragoza, España.

Msc., Universidad de Zaragoza, España.

Dr., Universidad de Zaragoza, España

Fecha de defensa de la tesis: 15 de Abril de 2013

A mis padres, María Cristina Viñals y José Antonio Suárez, por apoyarme en mis estudios de grado y postgrado, por su incondicional ayuda, por sus valiosos consejos, por sus ejemplos de perseverancia y constancia pero, más que nada, por su amor y cariño.

A mi gran amor, Oscar Fernando Rhodas, por acompañarme en esta etapa tan importante de mi vida.

Agradecimientos

La culminación de la presente tesis doctoral es el resultado de un extenso proceso de aprendizaje y trabajo, el cual fue posible por la colaboración y el apoyo de muchas personas e instituciones a las que quiero agradecer. En primer lugar, agradezco a mi director de tesis el Dr. Rodolfo “Fito” Cantet. Su presencia incondicional, sus apreciados y relevantes aportes, sus críticas, sus comentarios y sus sugerencias fueron fundamentales para el desarrollo de esta investigación. Sin su extensa experiencia, y su sobresaliente trayectoria en el área de mejoramiento genético animal, no hubiese podido llegar a esta instancia. ¡Gracias Fito! Fue un verdadero honor contar con su dirección. Mi mayor agradecimiento a usted por contribuir a constituirme académicamente y formarme en el campo de la investigación, área nueva para mí hasta el momento en que lo conocí.

Agradezco al Dr. Walter Sosa Escudero por su valiosa asistencia en el capítulo 4.

A los miembros del jurado, Dr. Daniel Omar Maizon (Argentina), Dr. Jorge Ignacio Urioste Aguerre (Uruguay) y Dr. Luis Varona Aguado (España), por tomarse el trabajo desinteresado de leer el manuscrito, y discutir sobre los resultados de esta investigación.

Al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por otorgarme una beca de postgrado, la cual me permitió financiar gran parte del doctorado.

A la Asociación Argentina de Brangus, y al Dr. Claudio Fioretti de la empresa Estancias y Cabaña Las Lilas S.A., Argentina, por facilitarme sus datos de campo.

A los docentes de la Maestría en “Biometría y Mejoramiento” de la Escuela para graduados “Alberto Soriano” que me acompañaron en mi formación doctoral. Principalmente a la Dra. María del Carmen Fabrizio, Msc. Nidia Nora Abbiati, Msc. Norberto Bartoloni, Msc. Ana María Pereyra y Msc. Ana Nélide Birchmeier. Gracias a ellos aprendí mucho de lo que hoy sé acerca de álgebra, estadística, genética y programación.

A mis compañeros de la cátedra de Mejoramiento Genético Animal de la FAUBA: Eduardo Cappa, Sebastián Munilla Leguizamón, Laura Pruzzo, Valeria Schindler, Mónica Santos Cristal, Natalia Forneris, Yeni Bernal Rubio, José Luis Gualdrón, Juan David Corrales y Carolina García Baccino, por toda la ayuda que me han prestado así como los buenos momentos que he pasado en su compañía.

A mis compañeros de clases y amigos, Fabricio Bongiorno, Pamela Sarco y Teresa Argibay, por hacer mi proceso de aprendizaje mucho más ameno y divertido.

Al Departamento de Producción Animal de la Facultad de Agronomía de la Universidad de Buenos Aires (FAUBA), por la disponibilidad de sus instalaciones y recursos.

A todo el personal de la Escuela para Graduados "Alberto Soriano" por su cordial atención.

Por último, agradezco el apoyo y el cariño que me brindan mis seres más queridos: mi familia y amigos.

A todos, sinceramente...¡Muchas gracias!

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en ésta u otra institución.

María José Suárez

Publicaciones derivadas de la tesis

Capítulo 2:

Suárez, M. J.; Cantet, R. J. C. 2011. Bayesian estimation of maternally ancestral animal models for weaning weight of beef cattle. *Journal of Animal Science* 89:3859–3871.

Suárez, M. J.; Cantet, R. J. C. 2010. Estimación paramétrica en modelos con efectos ambientales maternos no independientes en Angus. 33° Congreso Argentino de Producción Animal. *Rev. Arg. Prod. Anim.* 30 (Supl.1):85–86.

Cantet, R. J. C.; **Suárez, M. J.** 2009. Estimación bayesiana de matrices genéticas con distintos grados de credibilidad. 32° Congreso Argentino de Producción Animal. *Rev. Arg. Prod. Anim.* 29 (Supl.1):6–8.

Capítulo 3:

Suárez, M. J.; Birchmeier, A. N.; Cantet, R. J. C. 2012. On parsimonious and equivalent animal models with (grand) maternal effects and missing (grand) dams. *Livestock Science* 150:324–336.

Suárez, M. J.; Cantet, R. J. C. 2008. Especificación de la varianza aditiva cuando las abuelas maternas son desconocidas. 31° Congreso Argentino de Producción Animal. *Rev. Arg. Prod. Anim.* 27 (Supl.1):115–116.

ÍNDICE GENERAL

	<u>Página</u>
DEDICATORIA.....	IV
AGRADECIMIENTOS.....	V
DECLARACIÓN.....	VII
PUBLICACIONES DERIVADAS DE LA TESIS.....	IX
ÍNDICE GENERAL.....	XI
ÍNDICE DE CUADROS.....	XIII
ÍNDICE DE FIGURAS.....	XIV
ABREVIATURAS.....	XV
RESUMEN.....	XVII
ABSTRACT.....	XVIII
 CAPÍTULOS	
1. INTRODUCCION.....	1
 2. ESTIMACIÓN BAYESIANA EN MODELOS ANIMALES CON EFECTOS ANCESTRALES ENTRE MADRES E HIJAS PARA EL CARÁCTER PESO AL DESTETE, EN BOVINOS DE CARNE.....	6
2.1. Introducción.....	6
2.2. Materiales y Métodos.....	7
2.2.1. Datos.....	7
2.2.2. Modelos.....	8
2.2.3. Estimación Bayesiana de los componentes de (co)varianza.....	10
2.2.3.1. Distribuciones <i>a priori</i>	10
2.2.3.2. Distribuciones <i>a posteriori</i>	12
2.2.3.3. Muestreo de G_o utilizando distintos grados de credibilidad.....	14
2.2.3.4. Muestreo de λ	15
2.2.3.5. Esquema de muestreo para el algoritmo de GS en el modelo mixto.....	16
2.2.3.6. Muestreo de <i>Gibbs</i>	16
2.3. Resultados y Discusión.....	17
2.3.1. Efectos aditivos de abuela materna.....	22
2.3.2. Modelos con efectos ambientales.....	22
2.3.3. Modelos que combinan efectos genéticos de abuela materna y efectos ambientales.....	24
2.3.4. Correlaciones de Spearman entre los valores de cría predichos....	24
2.3.5. Recuperación de la variación aditiva total en los modelos con efectos maternos.....	25
2.3.6. Tópicos algorítmicos y de convergencia.....	25
2.3.7. General.....	25
 3. MODELOS ANIMALES CON EFECTOS MATERNOS O DE ABUELA MATERNA EQUIVALENTES, CUANDO LAS MADRES Y/O LAS ABUELAS MATERNAS SON DESCONOCIDAS.....	27
3.1. Introducción.....	27
3.2. Especificación de la covarianza aditiva para registros donde las madres son desconocidas.....	28

3.2.1. Ejemplo numérico.....	31
3.3. Especificación de la covarianza aditiva para registros donde las abuelas maternas son desconocidas.....	35
3.3.1. Obtención de la varianza de los ϵ_{oi}	36
3.3.2. Un pequeño ejemplo numérico.....	39
3.4. Discusión.....	43
 4. PREDICCIÓN DE LOS VALORES DE CRÍA DE ANIMALES NACIDOS POR TRANSPLANTE EMBRIONARIO, EN EL MODELO CON EFECTOS MATERNOS.....	44
4.1. Introducción.....	44
4.2. Metodología para atenuar el sesgo por edad de madre incierta.....	45
4.2.1 Modelos con error de medición.....	45
4.2.1.1. Modelo clásico de error de medición.....	46
4.2.1.2. Corrección por sesgo.....	48
4.2.1.3. Modelo con error de medición clásico en el efecto fijo edad de la madre.....	49
4.2.1.4. Efecto de la metodología MEM sobre la EM incierta en Brangus argentino.....	50
4.3. Metodología propuesta para considerar la covarianza residual.....	51
4.3.1. Suavización de la covarianza residual en rodeos con datos de transplantes embrionarios.....	51
4.3.2. Observaciones.....	52
4.3.3. <i>Splines</i> Penalizadas	53
4.3.3.1. <i>B-Splines</i> utilizando la covariable día de nacimiento.....	54
4.3.3.2. Algoritmo de Cox-De Boor.....	55
4.3.3.3. Penalizaciones.....	56
4.3.3.4. Suavizado de los residuales en rodeos Brangus argentinos	58
4.4. Resultados.....	61
4.5. Implementación en la evaluación genética.....	68
4.6. Discusión.....	71
 5. DISCUSIÓN GENERAL.....	73
 BIBLIOGRAFÍA CITADA	75

ÍNDICE DE CUADROS

<u>Cuadro</u>	<u>Página</u>
2.1. Descripción del pedigrí correspondiente al rodeo Angus.....	7
2.2. Medias marginales posteriores con sus respectivos desvíos estándares, e intervalos de alta densidad posterior (95% IADP, en paréntesis) para los parámetros de dispersión estimados en los ocho modelos propuestos.....	18
2.3. Coeficientes de correlación por rangos de Spearman calculados entre los valores de cría directos (encima de la diagonal) y, entre los valores de cría maternos (debajo de la diagonal) de los ocho modelos animales propuestos.....	24
4.1. Soluciones para el efecto EM para peso al nacer y al destete con dos modelos en bovinos de carne de la raza Brangus.	51
4.2. Descripción de los establecimientos de la raza Brangus.....	53
4.3. Extensión de los nodos por establecimiento.....	59
4.4. Estimaciones de los componentes de (co)varianza por cabaña.....	61

ÍNDICE DE FIGURAS

<u>Figura</u>	<u>Página</u>
2.1. Distribuciones marginales posteriores de la correlación entre los efectos ambientales maternos permanentes madre-hija (λ) de los modelos E, ESY, GE y GESY.....	20
2.2. Distribuciones marginales posteriores de la proporción de la variancia fenotípica que es debida a la interacción toro \times año (c_{SY}) en los modelos ISY, ESY, GSY y GESY.....	21
4.1. Representación esquemática de las relaciones funcionales entre funciones básicas B-splines de órdenes crecientes.....	55
4.2. Gráficos de Residuales vs Valores Predichos de los 12 establecimientos analizados.	62

Abreviaturas

ACF	Funciones de autocorrelación simple;
AIC	Criterio de información de Akaike;
AR(1)	Procesos autoregresivos de orden 1;
BLUP	Predicción lineal insesgada de mínima varianza;
BOA	Bayesian Output Analysis;
<i>B-splines</i>	<i>Splines</i> básicas;
DF	Prueba de Dickey Fuller;
DN	Día de nacimiento;
DS	Desvío estándar;
E	Modelo animal con efectos maternos con una matriz de covarianzas de tipo “autoregresivo” entre los efectos ambientales permanentes de madres emparentadas por vía materna;
EM	edad de madre;
ESY	Modelo animal con efectos maternos con una matriz de covarianzas de tipo “autoregresivo” entre los efectos ambientales permanentes de madres emparentadas por vía materna, y un efecto aleatorio de interacción toro \times año;
Fdp	Función de densidad posterior;
G	Modelo animal con efectos aditivos de abuela materna;
GE	Modelo animal con efectos aditivos de abuela materna, y con una matriz de covarianzas de tipo “autoregresivo” entre los efectos ambientales permanentes de madres emparentadas por vía materna;
GESY	Modelo igual a GE pero en este caso se tiene en cuenta, además, un efecto aleatorio de interacción toro \times año;
GS	Muestreo de Gibbs;
GSY	Modelo animal con efectos aditivos de abuela materna, y con un efecto aleatorio de interacción toro \times año;
I	Modelo animal con efectos maternos convencional;
ISY	Modelo animal con efectos maternos que incluye un efecto aleatorio de interacción toro \times año;
IADP95	Intervalo de alta densidad posterior al 95%;
MCMC	Cadenas de Markov Monte Carlo;

MEM	Modelos con errores de medición;
MME	Ecuaciones de modelo mixto;
MOET	Esquemas de ovulación múltiple y transferencia embrionaria;
PACF	Funciones de autocorrelación parcial;
PD	Pesos al destete;
PN	Pesos al nacer;
<i>P-splines</i>	<i>Splines</i> Penalizadas;
REML	Máxima Verosimilitud Restringida;
SAS	Statistical Analysis System;

RESUMEN

Modelos animales para caracteres con efectos maternos, madres no identificadas e información incompleta.

La evaluación genética para caracteres de crecimiento pre-destete requiere ajustar modelos animales con efectos maternos (MAM). Tanto la estimación paramétrica de la variabilidad como la evaluación genética mediante MAM son realizadas empleando datos de campo, muchos de los cuales no poseen información completa para todas las variables explicativas maternas. Es común no contar con la identificación de madres (biológicas y/o receptoras), de abuelas maternas y, consecuentemente, de la edad de la madre (EM). Este problema es bien marcado en razas compuestas como Brangus y Braford que tienen políticas para registrar animales de pedigrí “abierto”. Además, no existe un consenso sobre cuál es el mejor modelo de predicción, y existen interrogantes sobre la magnitud de los componentes de (co)varianza genético-aditivos y ambientales del modelo de evaluación. La primera investigación de esta tesis consistió en la estimación, mediante métodos bayesianos de los parámetros de dispersión en MAMs con distintas estructuras de (co)varianza, para datos de peso al destete de animales Angus de pedigrí. El análisis se caracterizó por la originalidad en los muestreos de las distribuciones marginales posteriores de las covarianzas genéticas aditivas y de la correlación entre los efectos ambientales maternos permanentes de una vaca y sus hijas también madres. Con el objeto de especificar correctamente la fracción aditiva de las (co)varianzas cuando se desconocen las madres y/o abuelas maternas de los animales con datos, en otro capítulo se desarrollaron MAMs equivalentes que no requieren alargar los vectores de los valores de cría con madres o abuelas fantasmas. Finalmente, se desarrolló un modelo mixto que atenúa el sesgo por error de medición clásico en el efecto EM, e introduce *splines* penalizadas y una estructura de (co)variación autoregresiva de orden 1 para suavizar las covarianzas residuales. Este modelo es apropiado para ajustar datos de animales nacidos por transplante embrionario con madres receptoras desconocidas.

Palabras Clave: Efectos maternos; efectos aditivos de abuela materna; madres desconocidas; abuelas maternas desconocidas; transplantes embrionarios; edad de madre incierta; modelos con errores de medición; *splines* penalizadas; estructuras autoregresivas de orden 1.

ABSTRACT

Animal models for maternally affected traits when dams are unknown and the information incomplete.

Maternal animal models (MAM) are used for the genetic evaluation of pre-weaning traits. Both the parametric estimation of variability and the genetic evaluation under MAM are performed using field data that often have maternal explanatory variables incompletely specified or unspecified. Therefore, lack of identification of the dam (either the biological, or the foster dam in case of embryo transplants), of the maternal grand-dam, and of the age of dam (EM), are quite common. The problem is relevant for the genetic evaluation of composites, such as Brangus or Braford, breeds that have “open” policies of animal registration. Additionally, there is no consensus on which model of prediction is best, and questions on the magnitude of the additive genetic covariance and environmental components in the genetic evaluation model arise. The first research chapter of the current thesis consisted of Bayesian estimation of dispersion parameters in MAM with different covariance structures, for weaning weights of pedigreed Angus calves. The originality of the analysis was in the algorithm used to sample the additive genetic (co)variances and the correlation between environmental effects of a cow and those of her daughters, which are also dams themselves. In another research chapter, equivalent MAMs were described for properly accounting for all additive genetic and environmental variances when dams or maternal grand-dams of animals with records are unknown. The procedure does not require enlarging the vector of breeding values with neither those of “phantom” dams, nor with those of “phantom” maternal grand-dams. Finally, a mixed animal model was proposed to alleviate the bias introduced classic measurement error in the effect age of dam, at the same time that penalized *splines* and an autoregressive covariance structure of order 1 are used to smooth residual covariances. The model is appropriate for fitting data for a maternally influenced trait, from embryo transfer calves with unknown foster dams.

Key words: maternal effects; grand maternal genetic effects; unknown dams; unknown grand dams; embryo transfers; unknown age of dam; measurement error models; penalized splines.

Capítulo 1

Introducción

La evaluación del mérito genético animal, o más apropiadamente el valor de cría, para caracteres de crecimiento y calidad de res, es la herramienta de mejoramiento genético que mayor impacto ha tenido sobre la producción de bovinos de carne. Los modelos de evaluación para crecimiento pre-destete incluyen efectos maternos. Si bien las primeras evaluaciones genéticas fueron realizadas a mediados de la década de 1970-1980, aún hoy no existe un consenso sobre cuál es el mejor modelo de predicción (Meyer, 1997; Bijma, 2006) y existen interrogantes sobre la magnitud de los componentes de (co)varianza genético aditivos y ambientales del modelo de evaluación. Dichos parámetros son la varianza aditiva directa (σ_{Ao}^2), la covarianza genética aditiva entre el valor de cría directo y el valor de cría materno (σ_{AoAm}), y las varianzas de los valores de cría (σ_{Am}^2) y ambiental (σ_{Em}^2) maternas. En muchas ocasiones, las bases de datos empleadas para estimar estas (co)varianzas directas y maternas no contienen suficiente información para generar estimaciones precisas de los parámetros de dispersión, de modo que puedan ser incluidas en la resolución de las ecuaciones de modelo mixto (Meyer, 1997). En particular, estimaciones altamente negativas de σ_{AoAm} conducen a predicciones con magnitudes contrapuestas de los valores de cría directo y materno. En este caso, animales con un alto mérito genético predicho para la componente directa tienden a mostrar valores de cría predichos por debajo del promedio para el valor de cría materno. Esto genera suspicacia y descrédito entre los criadores que utilizan las predicciones de la evaluación genética como herramienta de selección, dado que los animales que muestran mayor magnitud para la componente aditiva directa del carácter tienden a ser pobres reproductores para la aptitud materna. Si bien cierta asociación genética adversa entre efectos directos y maternos es aceptada entre los investigadores (Koch, 1972; Meyer, 1997; Bijma, 2006), correlaciones extremadamente negativas (por ejemplo, menores a -0.5) son consideradas con escepticismo tanto por los criadores como por los técnicos (Meyer, 1997).

Una característica importante del crecimiento pre-destete es que, tanto la estimación paramétrica de la variabilidad como la evaluación genética, son realizadas con datos de campo donde una fracción variable de los registros no muestra integridad de la información para todas las variables explicativas maternas: 1) la identificación de la madre; 2) la edad de la madre al momento de medición del carácter. La evaluación genética animal empleando modelos mixtos y predicción lineal insesgada de mínima varianza (BLUP, Henderson, 1984) se realiza condicional a que las variables explicativas consideradas como efectos fijos (sexo, edad del animal, edad de la madre, grupo de contemporáneos, etc) sean medidas sin error, y que la matriz de (co)varianzas de los efectos aleatorios (valores de cría y efectos ambientales permanentes) sean especificadas correctamente (Henderson, 1984). Sin embargo, es común que aparezcan variables explicativas observadas incompletamente. Por ejemplo, el 16% de los animales en la evaluación genética del Brangus argentino provienen de vacas con edad (EM) incierta, sean hembras receptoras de transferencia embrionaria o que gestan su propio ternero (Cantet, 2005). En consecuencia, se introduce un sesgo en los valores de cría predichos para los caracteres afectados en los datos de terneros hijos de dichas

madres. Por ejemplo, para peso al nacer y al destete, las predicciones del valor de cría de los hijos de vaquillonas y vacas viejas tienden a ser sesgados hacia abajo y, los de vacas adultas están sobre-predichos. El problema es aún más serio si las hembras de transplante embrionario (TE) son identificadas separadamente, debido a la imposibilidad de diferenciar el nivel genético superior de los animales de TE, usualmente selectos, con el verdadero efecto EM. Para remediar tal situación, existen metodologías estadísticas que permiten atenuar el sesgo cuando las variables explicativas son medidas con error (Fuller, 1987; Carroll, 2005; Buzas et al., 2005). Dichas técnicas son ampliamente difundidas en estudios epidemiológicos, y no han sido aplicadas en problemas de evaluación genética animal.

Adicionalmente, en ganado de carne existen ciertas razas o sub-poblaciones con políticas “abiertas” para registrar animales, a diferencia de lo que ocurre con las razas británicas tradicionales como Angus o Hereford, donde la inscripción al libro genealógico (herdbook) sólo puede producirse cuando ambos padres del animal se encuentran ya registrados. En las razas de pedigrí “abierto” como Brangus o Braford, una fracción importante de animales con datos provienen de terneros cuyas madres no tienen medición propia para los caracteres maternalmente afectados y, además, ninguno de sus padres (los abuelos maternos del ternero con dato) se encuentra identificado. Como consecuencia de dicho proceso, la evaluación genética de razas abiertas para caracteres con efectos maternos sufre de errores de especificación dado que, cuando un individuo con datos posee información faltante en relación con la madre y/o su abuela materna (Cantet et al., 1992b), la especificación de la estructura de covarianzas del modelo es distinta de la covarianza entre parientes con efectos maternos que postulara Willham (1963) y que es la especificada en el modelo animal con efectos maternos de Quaas y Pollak (1980). En tales condiciones, los predictores de los valores de cría resultantes no son de mínima varianza (Henderson, 1975). Cuando una fracción de las madres de los individuos con datos no está identificada, Henderson (1988) observó que en la varianza del dato individual, el término del error del modelo absorbe los efectos genéticos y ambientales maternos permanentes y, por lo tanto, la varianza del modelo no es homoscedástica. Una alternativa frente a este problema es aumentar el vector de valores de cría agregándole las madres desconocidas o “fantasmas” (Cantet et al., 1992b) y sumar σ_{Em}^2 a la varianza del error. Como cada una de las madres fantasmas tiene una sola cría, la matriz de covarianzas del error es diagonal (Cantet et al., 1992b). Un problema similar ocurre con las madres receptoras de los animales nacidos por transplante embrionario las cuales no son reportadas, con el agravante de que suelen estar repetidas entre los registros de un mismo establecimiento a lo largo del tiempo. En esta última situación es imposible especificar correctamente la matriz de (co)varianzas residuales, aún cuando se incluya en el vector de valores de cría a los de las madres receptoras fantasmas. La consecuencia de esta sub-especificación es la presencia de covarianza residual en los datos de TE de un establecimiento. Frente a la imposibilidad de especificar la verdadera estructura de (co)varianzas, una alternativa es realizar un “suavizado” de los residuales para descomponerlos en una porción correlacionada y otra homoscedástica, como se desprende del trabajo fundacional de Heckman (1986). Estas técnicas semiparamétricas que permiten relajar el supuesto restrictivo acerca de la forma funcional conocida, se pueden implementar mediante modelos mixtos y *splines* penalizadas (ver el texto de Ruppert et al., 2003). Esta formulación fue utilizada en un contexto de evaluación genética animal con especificación de grupos de contemporáneos incierta por Cantet et al. (2005) y, en un contexto de evaluación

genética forestal con residuales correlacionados espacialmente por Cappa y Cantet (2007) y amerita ser analizada para el caso que se describe en este párrafo.

Existen otros modelos alternativos que atenúan el valor estimado negativo del parámetro σ_{AoAm} , descomponiendo la covarianzas ancestrales entre madres e hijas dentro de la covariación genético-ambiental materna. Una parametrización relativamente parsimoniosa, pero exigente en cuanto a requerimientos de número de terneros por vaca y relaciones de parentesco entre las madres, es la de Quintanilla et al. (1999). Estos autores propusieron una estructura de covarianza correlacionada por un proceso autoregresivo para los efectos ambientales maternos y lograron reducir el valor negativo estimado de σ_{AoAm} en una base de datos reales.

Otro modelo con efectos maternos empleado para atenuar el sesgo y la precisión de las estimaciones de σ_{AoAm} incluye las interacciones toro \times rodeo, o toro \times año, o toro \times grupo de contemporáneos (Robinson, 1996; Lee y Pollak, 1997; Gutiérrez et al., 2006). La inclusión de alguna de dichas interacciones resultó en un valor estimado de σ_{AoAm} con magnitud menos negativa que cuando dicha interacción era excluida del modelo. La reducción del valor negativo del parámetro fue variable según el trabajo pero consistente en dirección: siempre la magnitud negativa de σ_{AoAm} disminuyó en los trabajos de Robinson (1996); Lee y Pollak (1997) y Gutiérrez et al. (2006). Otro modelo con efectos maternos alternativo empleado para obtener estimaciones menos negativas de σ_{AoAm} incluye los valores de cría por la vía ancestral de la abuela materna (Willham, 1972). Este modelo es sobreparametrizado con tres componentes de covarianza aditivos, adicionales a los tres del modelo con efectos maternos: la varianza de los valores de cría de abuela materna σ_{An}^2 , la covarianza entre los valores de cría directos y de abuela materna σ_{AoAn} y la covarianza entre los valores de cría maternos y los de abuela materna σ_{AmAn} (Willham, 1972). La estimación en el modelo de abuela materna es inherentemente compleja, por lo tanto su aplicación es más reciente (Dodenhoff et al., 1998, 1999a,b). No existen resultados concluyentes sobre la magnitud estimada de los efectos de abuela materna en las estimaciones de σ_{AoAm} . Por ejemplo, los valores estimados inicialmente por Dodenhoff et al. (1998) son optimistas, en comparación con aquellas obtenidas posteriormente (Dodenhoff et al., 1999b). Aparentemente, una posible diferencia entre las estimaciones se deba a la integridad de la identificación de las madres y abuelas maternas en las distintas bases (Dodenhoff et al., 1999b). Asimismo, parecería que la magnitud de σ_{Am}^2 es subestimada y que σ_{AoAm} es menos negativa cuando se excluyen los efectos de abuela materna del modelo (Dodenhoff et al., 1999b). Es dable esperar diferencias importantes en la precisión de las estimaciones de los componentes de (co)varianza en modelos con efectos maternos y de abuela materna, originadas en la cantidad de información presente en los datos y la integridad del pedigrí (Van Vleck, 1990a; Cantet et al., 1992b; Dodenhoff et al., 1999a,b, Gutiérrez et al., 2006). Además, sería problemático seleccionar los “grados de credibilidad” de la distribución Wishart invertida (Sorensen y Gianola, 2002, página 574), si se empleasen métodos Bayesianos de estimación, dado que existe muy escasa información previa para la varianza aditiva de abuela materna en comparación con σ_{Am}^2 , y más aún en relación con lo que se conoce sobre la magnitud de σ_{Ao} .

En virtud de todo lo expuesto precedentemente, los objetivos generales de la presente tesis son los siguientes:

- 1) Estimar mediante métodos bayesianos los parámetros de dispersión de modelos animales con efectos maternos empleando diferentes estructuras de (co)varianza;

- 2) Desarrollar modelos animales con efectos maternos que especifiquen completa y correctamente la estructura de (co)varianzas aditivas y ambientales, cuando las madres y las abuelas maternas de los individuos con datos no están identificadas o la información de las hembras sea incompleta (ejemplo, la edad de la madre);

Para cumplir con estos objetivos, la tesis está organizada en cinco capítulos, incluyendo el presente de naturaleza introductoria. En el segundo de ellos se describen siete formulaciones alternativas al modelo animal con efectos maternos clásico (Willham, 1963; Quaas y Pollak, 1980) para el peso al destete, combinando efectos ambientales maternos permanentes correlacionados, efectos aditivos de abuela materna y/o una interacción toro \times año. Luego se describen los algoritmos Bayesianos MCMC que permiten estimar los parámetros de dispersión de todos los modelos presentados. Para el muestreo de la distribución marginal posterior de la covarianza genética aditiva se describe una distribución conjugada Wishart Invertida Generalizada (Brown, 2001). A diferencia de la distribución Wishart Invertida usualmente empleada en modelos con efectos maternos o multicarácter (Sorensen y Gianola, 2002), la GIW permite utilizar distintos “grados de credibilidad” para los diferentes componentes de (co)varianza (Cantet et al., 2004; Munilla y Cantet, 2012), es decir, distinta confiabilidad de la información *a priori* y *a posteriori*, dado su naturaleza conjugada. Adicionalmente, se describe el muestreo de la correlación entre los efectos ambientales maternos permanentes de una madre y sus hijas, que a su vez son madres, a partir de una distribución normal truncada siguiendo los lineamientos de Chib (1993). Este método es distinto al empleado por Quintanilla et al. (1999), quienes utilizaron el algoritmo Metrópolis-Hastings por etapas dentro del muestreo de Gibbs. Los ocho modelos presentados se ajustaron a una base de datos informativa de un rodeo puro de pedigrí de la raza Angus. El capítulo finaliza con el análisis del impacto que tiene la inclusión de los efectos maternos ancestrales y la interacción toro \times año sobre la magnitud de r_{AoAm} , (es decir, σ_{AoAm} expresada como correlación aditiva), la heredabilidad total (Willham, 1972) y el cambio en el orden del mérito genético de los animales en evaluación.

En el capítulo 3, se proponen dos modelos con efectos maternos y de abuela materna equivalentes a los presentados por Willham (1972), Cantet et al. (1992b) y Van Vleck (1990a), cuando las madres o las abuelas maternas son desconocidas. A tal efecto, se emplea un enfoque de “regresión” que fuera originalmente propuesto por R. L. Quaas para la revisión del trabajo de Cantet et al. (1992b). Esta solución no requiere agrandar el orden de las ecuaciones a resolver, a diferencia de los modelos que incluyen madres o abuelas fantasmas (Van Vleck, 1990a). Finalmente, los modelos propuestos se ejemplifican mediante dos ejemplos numéricos.

En el capítulo 4 se presentan dos soluciones al problema de la evaluación genética cuando en los datos de los animales nacidos por trasplante embrionario no se encuentran identificadas las madres receptoras. En primer lugar, se aborda el problema de la edad de madre incierta mediante la inclusión de errores de medición clásicos en el efecto fijo edad de la madre (EM). Se muestra cómo corregir las ecuaciones de modelo mixto cuando se presentan errores de medición aleatorios, es decir, en cualquier

categoría del efecto fijo, y se conocen tanto la varianza del error de medición, como la varianza de la variable bajo estudio medida sin error. Luego, la metodología es aplicada al análisis de EM para el peso al nacer y al destete de vacunos Brangus de Argentina. Después, se cuantifica la magnitud de la (co)varianza presente en los residuales del peso al destete para los animales pertenecientes a rodeos con registros de trasplantes embrionarios. Para ello, se emplean técnicas de series de tiempo equiespaciadas a través del ajuste de *splines* penalizadas (Cantet et al., 2005). Este enfoque facilita la utilización de la metodología de Box y Jenkins (1976) para, posteriormente, implementar un proceso AR(1) tal como lo proponen Wade y Quaas (1993). Este último procedimiento simplifica el cálculo para la inversión de la matriz de covarianzas de una gran base de datos, y posibilita la resolución de las ecuaciones de modelo mixto para la evaluación genética.

El capítulo 5 contiene las conclusiones generales de la investigación descripta en los capítulos 2, 3 y 4. A su vez, se señala el significado de los resultados obtenidos, y las futuras líneas de investigación.

En lo que respecta a la nomenclatura utilizada a lo largo del manuscrito, se respetaron las siguientes reglas:

- Las matrices se representan mediante letras mayúsculas del alfabeto castellano en negrita e itálica o mayúsculas del alfabeto griego en negrita;
- Los vectores se denotan con letras minúsculas del alfabeto castellano en negrita, o en letras del alfabeto griego en negrita;
- Las letras minúsculas tanto del alfabeto castellano como del griego sin negrita y en cursiva representan escalares;
- I representa a la matriz identidad;
- 0 es el vector con todos sus elementos iguales a cero;
- 1 es un vector con todos sus elementos iguales a uno;

Capítulo 2

Estimación Bayesiana en modelos animales con efectos ancestrales entre madres e hijas para el carácter peso al destete, en bovinos de carne.

2.1. INTRODUCCIÓN

A pesar de que las evaluaciones genéticas para caracteres de crecimiento predestete en bovinos de carne fueron realizadas a mediados de la década de 1970-1980, aún no existe un consenso sobre cuál es la mejor parametrización de los componentes de (co)varianza (Bijma, 2006; Munilla Leguizamón y Cantet, 2010), o cómo modelar el efecto del ambiente sobre la magnitud de los componentes de (co)varianza genéticos aditivos (Speidel et al., 2007), dentro del modelo animal con efectos maternos. El aspecto más problemático es que la presencia de estimaciones altamente negativas de la correlación genética directa - materna (r_{AoAm}) conducen a predicciones contrapuestas de los valores de cría de cada componente. En tal situación, animales con un alto mérito genético predicho para la componente directa tienden a presentar valores de cría predichos por debajo de la media para la expresión materna del carácter. Esto genera suspicacia y descrédito entre los criadores que utilizan las predicciones de la evaluación genética como herramienta de selección, dado que los animales que muestran mayor magnitud de la componente directa tienden a ser pobres reproductores en cuanto a aptitud materna (Willham, 1972; VanVleck et al., 1987; Meyer, 1997).

Se han propuesto distintos modelos teóricos para explicar la covarianza entre el registro del peso al destete de una vaca y el de su progenie: 1) la inclusión de un efecto de abuela materna (Willham, 1972; Doddenhoff et al., 1999a); 2) la consideración de una correlación entre los efectos ambientales maternos permanentes de una madre y sus hijas (Quintanilla et al., 1999); 3) la inclusión de un efecto aleatorio de interacción toro \times año o toro \times rodeo (Robinson, 1996; Lee y Pollak, 1997), o toro \times grupo de contemporáneos (Gutiérrez et al., 2006). Sin embargo, no hay ningún antecedente en la literatura sobre cómo estos modelos animales con efectos maternos redistribuyen la variabilidad aditiva total y la ambiental entre sus parámetros de dispersión. La sobreparametrización es complicada por el hecho que, en muchas ocasiones, las bases de datos empleadas no contienen suficiente información para estimar todos los parámetros de dispersión (Meyer, 1997). Por lo tanto, el objetivo de este capítulo fue evaluar en qué medida los modelos de evaluación genética con efectos aditivos de abuela materna (G), o con una matriz de covarianzas de tipo “autoregresivo” entre los efectos ambientales permanentes de madres emparentadas por vía materna (E), o con una interacción toro \times año (ISY), o con una combinación de ellos (GE, GSY, ESY, GESY), redistribuyen la variabilidad aditiva total y reducen la magnitud negativa de r_{AoAm} , cuando se los compara con el modelo animal materno convencional (I). Para ello en este capítulo se describen algoritmos Bayesianos MCMC que permitirán estimar los parámetros de dispersión de los ocho modelos maternos presentados, en una informativa base de datos con terneros Angus.

2.2. MATERIALES Y MÉTODOS

2.2.1. Datos

Para la estimación de los componentes de (co)varianza de los modelos propuestos, se utilizaron los pesos al destete (PD) de terneros Angus de pedigrí, provenientes de un rodeo en la localidad de Pasteur, noroeste de la provincia de Buenos Aires, Argentina. Este archivo estuvo disponible en virtud del convenio marco vigente entre la empresa Estancias y Cabaña Las Lilas y la Facultad de Agronomía de la Universidad de Buenos Aires (FAUBA). Para el análisis, se utilizaron 7229 (*n*) registros de PD, medidos en animales nacidos entre los años 1974 a 2008, y una genealogía de 9936 (*q*) animales registrados, hijos de 747 padres y 3404 madres. Hubo además 1905 abuelas maternas identificadas. Este punto fue vital a la hora de elegir el archivo de datos debido a la necesidad de contar con relaciones de parentesco informativas, fundamentalmente por vía materna, para estimar la compleja estructura de dispersión de los modelos. La correcta especificación de la estructura de covarianzas para cualquier individuo con dato requiere que madre (Van Vleck, 1990a; Cantet et al., 1992b) y abuela materna (Van Vleck, 1990a; Doddenhoff et al., 1999b) sean conocidas. De otro modo, el modelo ajustado no explica totalmente la variación aditiva (Cantet et al., 1992b). El número de vacas con ternero con dato propio (*d*) fue 2444. Dado la compleja partición de la varianza aditiva y ambiental en animales criados con madres receptoras, se descartaron aquellos registros provenientes de animales nacidos por trasplante embrionario, como así también los partos de mellizos. El PD promedio de todos los datos fue $205,3 \pm 40,3$ kg y la edad promedio 194 ± 23 días. En el Cuadro 2.1 se muestra una descripción detallada de la estructura del pedigrí. Para conocer en detalle el manejo y procedimientos de selección del rodeo se pudo consultar Cantet et al. (1993, 2005) y Pruzzo et al. (2003).

Cuadro 2.1. Descripción del pedigrí correspondiente al rodeo Angus.

Pedigrí		Padres	Madres	Abuelas	Total
Nº de Animales		747	3404	1905	9936
Nº de Progenitores c/ registros de PD		55	1389	607	1444
% de Progenitores c/ registros de PD		7,36	40,80	31,86	14,53
Nº promedio de crías por progenitor		12,54	2,61	4,28	
% de progenitores con:	1 cría	31,05	43,00	29,29	
	2 crías	20,08	18,74	13,49	
	3 crías	9,23	12,37	12,02	
	4 crías	4,56	9,06	10,92	
	5 crías	4,69	6,20	7,35	
	> 5 crías	30,39	10,63	26,93	

PD = peso al destete; DS = desvío estándar.

2.2.2. Modelos

Los datos se analizaron con ocho modelos animales distintos, los cuales difieren únicamente en los efectos aleatorios a ajustar y/o en los supuestos que se asumen de ellos. Los efectos fijos que se tuvieron en cuenta fueron el sexo de los animales (machos, hembras), la edad de la madre (en 5 clases: 2, 3, 4, 5 a 8 y, 9 o más), los grupos de contemporáneos (96 niveles) y la edad del ternero, como una covariable lineal. Para que la matriz de incidencia (X) de los efectos fijos sea de rango completo, las soluciones de sexo de las hembra y el último nivel de la edad de madre se igualaron a cero; de esta manera, rango $[X] = p = 102$. En todos los modelos se tuvieron en cuenta los valores de cría directos (a_o) y maternos (a_m), los efectos ambientales maternos permanentes (e_m), además del término de error (e_o), como efectos aleatorios. A los modelos que además se le incluyeron los efectos genéticos de abuela materna (a_n , Willham, 1972) se los identificó como G. Se consideraron dos especificaciones de la matriz de covarianzas para el vector e_m : a) la estructura clásica con efectos ambientales maternos permanentes independientes entre sí (modelos I); b) la estructura propuesta por Quintanilla et al. (1999), la cual considera una correlación (λ) entre los efectos ambientales maternos permanentes de una madre y sus hijas, que a su vez son madres (modelos E). Para completar el análisis, se tuvo en cuenta, además, la interacción toro \times año (h , Robinson, 1996) identificada con la sigla SY, la cual contó con 1062 (s) niveles en este archivo de datos. Para simplificar la presentación, se presentará el modelo más general (GESY), que abarca todos los modelos restantes. La ecuación matricial en GESY es igual a:

$$y = X \beta + Z_o a_o + Z_m a_m + Z_n a_n + Z_p e_m + Z_{sy} h + e_o \quad [2.1]$$

donde y es un vector de orden 7229×1 con los registros fenotípicos observados; X es de orden 7229×102 ; β es el vector (102×1) de efectos fijos. Las matrices Z_o , Z_m y Z_n (todas de orden 7229×9936) respectivamente relacionan a_o , a_m , y a_n con y . La matriz Z_p (7229×2444) relaciona a los efectos ambientales maternos permanentes e_m (2444×1) con y ; Z_{sy} es la matriz de incidencia (7229×1062) que relaciona a h (1062×1) con y . Finalmente, e_o (7229×1) es el vector aleatorio de errores. Asumiendo que todas las relaciones de parentesco son conocidas, las variables aleatorias de la derecha de [2.1] tienen esperanza y estructura de varianzas y covarianzas iguales a:

$$E \begin{bmatrix} a_o \\ a_m \\ a_n \\ e_m \\ h \\ e_o \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \quad \text{Var} \begin{bmatrix} a_o \\ a_m \\ a_n \\ e_m \\ h \\ e_o \end{bmatrix} = \begin{bmatrix} A\sigma_{Ao}^2 & A\sigma_{AoAm} & A\sigma_{AoAn} & 0 & 0 & 0 \\ A\sigma_{AoAm} & A\sigma_{Am}^2 & A\sigma_{AmAn} & 0 & 0 & 0 \\ A\sigma_{AoAn} & A\sigma_{AmAn} & A\sigma_{An}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & E_d \sigma_{Em}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_s \sigma_{SY}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_n \sigma_{Eo}^2 \end{bmatrix} \quad [2.2]$$

donde A , de orden (9936×9936), es simétrica, positiva definida y es conocida como la matriz de relaciones aditivas porque contiene las ‘relaciones de parentesco’ entre los animales a ser evaluados. Henderson (1976) desarrolló un algoritmo para calcular su inversa directamente, sin nunca tener que generar A , el cual permite considerar la consanguinidad. Sin embargo, en este trabajo se utilizó el algoritmo propuesto por Quaas (1976). Los parámetros de dispersión a estimar fueron la varianzas aditivas directa (σ_{Ao}^2), materna (σ_{Am}^2) y de abuela materna (σ_{An}^2); las covarianzas genéticas

entre la componente directa y materna (σ_{AoAm}), directa y de abuela materna (σ_{AoAn}), materna y de abuela materna (σ_{AmAn}); la varianza de los efectos ambientales maternos permanentes (σ_{Em}^2); la correlación entre los efectos ambientales maternos permanentes de una madre y sus hijas (λ); la varianza de la interacción toro \times año (σ_{SY}^2) y, por último, la varianza del error (σ_{Eo}^2). Ahora bien, si escribimos $\mathbf{a} = [\mathbf{a}'_o, \mathbf{a}'_m, \mathbf{a}'_n]'$, y

definimos $\mathbf{G}_o = \begin{bmatrix} \sigma_{Ao}^2 & \sigma_{AoAm} & \sigma_{AoAn} \\ \sigma_{AoAm} & \sigma_{Am}^2 & \sigma_{AmAn} \\ \sigma_{AoAn} & \sigma_{AmAn} & \sigma_{An}^2 \end{bmatrix}$ como la matriz de covarianzas aditivas, tenemos

entonces que $\text{Var}(\mathbf{a}) = \mathbf{G}_o \otimes \mathbf{A}$, donde el símbolo \otimes representa al producto Kronecker (o producto directo). Esta estructura Kronecker de la matriz de covarianzas aditivas (\mathbf{G}_o) es la que facilita su inversión, dado que permite utilizar la fórmula explícita de la inversa de la matriz \mathbf{A} (Henderson, 1976; Quaas, 1976), y ha sido la clave de la amplia difusión de los modelos animales con efectos maternos entre los mejoradores animales.

Siguiendo la presentación de Quintanilla et al. (1999), $\mathbf{E} = \{f(\lambda)\}$, tal que $|\lambda| < 1$, es una matriz ($d \times d$) de correlaciones entre efectos ambientales maternos permanentes contribuidos por las madres de los individuos con datos y, λ es un nuevo parámetro que refleja la correlación entre los efectos ambientales maternos permanentes de una madre y sus hijas (con crías en el archivo) que, se asume constante a través de las generaciones. La matriz \mathbf{E} posee elementos diagonales iguales a 1 y, no-diagonales iguales a λ elevado a una potencia por generación, o cero en caso de vacas no relacionadas. De este modo, el parámetro λ tiene cierta “memoria” dado que rescata información de las nietas, bisnietas, y de todas las descendientes de una hembra que llegan a ser madres. Por ejemplo, sea i la madre de j que, a su vez es madre. Entonces, la $\text{Cov}(\mathbf{e}_{mi}, \mathbf{e}_{mj}) = \lambda \sigma_{Em}^2$. A su vez, cuanto menor es la magnitud de λ , menor es el impacto que tendrá la información más distante (Quintanilla et al., 1999). Alternativamente, en el modelo GSY los efectos ambientales maternos permanentes se asumieron independientes, es decir $\text{Var}(\mathbf{e}_m) = \mathbf{I}_{2444} \sigma_{Em}^2$. En el modelo ESY se asumió $\text{Var}(\mathbf{e}_m) = \mathbf{E} \sigma_{Em}^2$, se tuvo en cuenta la interacción toro \times año pero se excluyeron los efectos genéticos de abuela materna. En el modelo ISY se consideraron, tanto a los efectos ambientales maternos permanentes como a la interacción toro \times año, como niveles independientes. Por último, los modelos GE, G, E y I son idénticos a GESY, GSY, ESY y ISY con la excepción que no se consideró a la interacción toro \times año.

Cabe destacar que, en contraste con lo que ocurre con los modelos con efectos genéticos de abuela materna (G, GE y GESY), los modelos que incluyen únicamente efectos maternos (I, E, ISY y ESY) poseen en el vector \mathbf{a} dos efectos aleatorios en lugar de tres. Como consecuencia:

$$\mathbf{G}_o = \begin{bmatrix} \sigma_{Ao}^2 & \sigma_{AoAm} \\ \sigma_{AoAm} & \sigma_{Am}^2 \end{bmatrix} \quad [2.3]$$

2.2.3. Estimación Bayesiana de los componentes de (co)varianza

Los parámetros de dispersión se estimaron mediante un enfoque Bayesiano conjugado de modo que, las densidades *a priori* y *a posteriori* pertenezcan a la misma familia de distribuciones (Robert y Casella 1999; Cantet et al., 2004). Como en la sección anterior, los desarrollos que se presentan en la presente sección serán ejemplificados con el modelo GESY. Además, definiremos a $\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{a}', \boldsymbol{e}_m', \boldsymbol{h}']'$ como el vector de parámetros para dicho modelo. Desde el enfoque Bayesiano, la función de verosimilitud de un modelo lineal mixto es la función de densidad de probabilidad (*fdp*) del vector de observaciones (\mathbf{y}) condicional en los parámetros de posición ($\boldsymbol{\theta}$) y en la varianza del error (σ_{Eo}^2). También es necesario definir las distribuciones a priori de todos los parámetros. Para el peso al destete la *fdp* es normal n -variada:

$$\mathbf{y} | \boldsymbol{\theta}, \sigma_{\text{Eo}}^2 \propto N_n(\mathbf{W}\boldsymbol{\theta}, \mathbf{I}_n \sigma_{\text{Eo}}^2) \quad [2.4]$$

En [2.4], $\mathbf{W} = [\mathbf{X} | \mathbf{Z}_o | \mathbf{Z}_m | \mathbf{Z}_n | \mathbf{Z}_p | \mathbf{Z}_{\text{sy}}]$. Sin embargo, \mathbf{W} se puede adaptar a cada uno de los modelos propuestos recordando que en no todos se incluyen los efectos genéticos de abuela materna (\mathbf{Z}_n) ni la interacción toro \times año (\mathbf{Z}_{sy}). Ahora bien, la verosimilitud de los datos es proporcional a:

$$p(\mathbf{y} | \boldsymbol{\theta}, \mathbf{G}_o, \lambda, \sigma_{\text{Em}}^2, \sigma_{\text{SY}}^2, \sigma_{\text{Eo}}^2) \propto (\sigma_{\text{Eo}}^2)^{-\left(\frac{n}{2}\right)} \exp\left\{-\frac{(\mathbf{y} - \mathbf{W}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{W}\boldsymbol{\theta})}{2\sigma_{\text{Eo}}^2}\right\} \quad [2.5]$$

2.2.3.1. Distribuciones a priori

Para el vector de efectos fijos se asumió, *a priori*, una distribución normal p -variada con una varianza de gran magnitud, de modo de reflejar un elevado grado de incertidumbre previo a la observación de los datos. Se consideró que todos los elementos del vector $\boldsymbol{\beta}$ son variables aleatorias independientes e idénticamente distribuidas. Por lo tanto, $\boldsymbol{\beta} \sim N_p(\mathbf{O}, \mathbf{K})$ donde los elementos de la matriz diagonal \mathbf{K} son de magnitud considerable ($k_{ii} \geq 10^8$):

$$p(\boldsymbol{\beta} | \mathbf{K}) \propto \left| \prod_{i=1}^p k_{ii} \right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^p \left(\frac{\beta_i^2}{k_{ii}} \right)\right\} \quad [2.6]$$

Si bien en algunos trabajos se ha propuesto utilizar la distribución uniforme (véase, por ejemplo, Cantet et al., (1992a), Hobert y Casella (1996) demostraron que este supuesto puede conducir a distribuciones posteriores impropias. Por su parte, de acuerdo a la teoría genético-cuantitativa (Bulmer, 1985), la distribución conjunta de los efectos genéticos aditivos para los q animales es $\mathbf{a} | \mathbf{A}, \mathbf{G}_o \sim N_{lq}(\mathbf{O}, \mathbf{G}_o \otimes \mathbf{A})$, donde l es la dimensión de \mathbf{G}_o (3 en este caso). Por lo tanto, su densidad es proporcional a:

$$p(\mathbf{a} | \mathbf{A}, \mathbf{G}_o) \propto |\mathbf{G}_o|^{-\frac{q}{2}} \exp\left\{-\frac{1}{2} \mathbf{a}'(\mathbf{G}_o^{-1} \otimes \mathbf{A}^{-1})\mathbf{a}\right\} \quad [2.7]$$

A priori la matriz de covarianzas aditivas (\mathbf{G}_o) sigue una densidad Wishart invertida (IW): $\mathbf{G}_o \sim \text{IW}(\mathbf{G}_o^*, \nu_g)$, siendo \mathbf{G}_o^* la matriz de hipercovarianzas de orden $(l \times l)$ y ν_g los grados de credibilidad (Sorensen y Gianola 2002, pág. 57). Más formalmente:

$$p(\mathbf{G}_o | \mathbf{G}_o^*, \mathbf{v}_g) \propto |\mathbf{G}_o|^{-\frac{1}{2}(\mathbf{v}_g + I + 1)} \exp\left\{-\frac{1}{2} \text{tr}[\mathbf{G}_o^{*-1} \mathbf{G}_o^{-1}]\right\} \quad [2.8]$$

A *priori* se asume para el vector de efectos ambientales maternos permanentes una distribución normal d -variada, con matriz de (co)varianzas $\mathbf{E} \sigma_{\text{Em}}^2$ (Quintanilla et al., 1999) tal que $\mathbf{e}_m | \sigma_{\text{Em}}^2, \lambda \sim N_d(\mathbf{0}, \mathbf{E} \sigma_{\text{Em}}^2)$, o sea que

$$p(\mathbf{e}_m | \mathbf{E}, \lambda, \sigma_{\text{Em}}^2) \propto |\mathbf{E}|^{-\frac{d}{2}} \exp\left\{-\frac{1}{2} \mathbf{e}_m' \mathbf{E}^{-1} \mathbf{e}_m\right\} \quad [2.9]$$

Asimismo, *a priori* se asume para el vector de interacción toro \times año una distribución normal s -variada, con matriz de (co)varianzas $\mathbf{I} \sigma_{\text{SY}}^2$, es decir: $\mathbf{h} | \sigma_{\text{SY}}^2 \sim N_s(\mathbf{0}, \mathbf{I} \sigma_{\text{SY}}^2)$. Por lo tanto:

$$p(\mathbf{h} | \mathbf{I}, \sigma_{\text{SY}}^2) \propto (\sigma_{\text{SY}}^2)^{-\frac{s}{2}} \exp\left\{-\frac{1}{2 \sigma_{\text{SY}}^2} \mathbf{h}' \mathbf{h}\right\} \quad [2.10]$$

Los vectores β , \mathbf{a} , \mathbf{e}_m y \mathbf{h} se suponen, *a priori*, independientes entre sí.

Las varianzas $\sigma_{\text{Em}}^2, \sigma_{\text{SY}}^2$ y σ_{Eo}^2 siguen *a priori* densidades Chi-cuadrado invertidas (χ^2) escaladas e independientes. Consecuentemente tenemos que:

$$p(\sigma_{\text{Em}}^2 | \mathbf{v}_{\text{Em}}, s_{\text{Em}}^2) \propto (\sigma_{\text{Em}}^2)^{-\left(\frac{\mathbf{v}_{\text{Em}}}{2} + 1\right)} \exp\left\{-\frac{\mathbf{v}_{\text{Em}} s_{\text{Em}}^2}{2 \sigma_{\text{Em}}^2}\right\} \quad [2.11]$$

$$p(\sigma_{\text{SY}}^2 | \mathbf{v}_{\text{SY}}, s_{\text{SY}}^2) \propto (\sigma_{\text{SY}}^2)^{-\left(\frac{\mathbf{v}_{\text{SY}}}{2} + 1\right)} \exp\left\{-\frac{\mathbf{v}_{\text{SY}} s_{\text{SY}}^2}{2 \sigma_{\text{SY}}^2}\right\} \quad [2.12]$$

$$p(\sigma_{\text{Eo}}^2 | \mathbf{v}_{\text{Eo}}, s_{\text{Eo}}^2) \propto (\sigma_{\text{Eo}}^2)^{-\left(\frac{\mathbf{v}_{\text{Eo}}}{2} + 1\right)} \exp\left\{-\frac{\mathbf{v}_{\text{Eo}} s_{\text{Eo}}^2}{2 \sigma_{\text{Eo}}^2}\right\} \quad [2.13]$$

En [2.11], [2.12] y [2.13], los símbolos s^2 y \mathbf{v} con sus respectivos subíndices representan a las *hipervarianzas* y a los *grados de credibilidad*, ambos parámetros de la Chi-cuadrado invertida. Finalmente, para λ se asumió *a priori* una distribución normal truncada, de modo tal que $-1 < \lambda < 1$ con una media *a priori* de -0.19 tal como estimaron Quintanilla et al. (1999).

2.2.3.2. Distribuciones *a posteriori*

Para la estimación de los componentes de (co)varianza de los modelos propuestos se utilizó el muestreo de Gibbs. Este algoritmo requiere las expresiones exactas de las densidades condicionales posteriores de los parámetros a estimar, de las que se muestrea en forma iterativa hasta obtener las distribuciones marginales (ver, por ejemplo, el texto de Sorensen y Gianola, 2002). En este apartado, la atención se va a centrar en la presentación de las distribuciones *a posteriori* de los parámetros del modelo GESY.

La distribución posterior conjunta de todos los parámetros es proporcional a la verosimilitud y al producto de las *fdp* correspondientes a las distribuciones a priori definidas previamente (Zellner, 1971; Box y Tiao, 1973). En este caso:

$$p(\boldsymbol{\theta}, \mathbf{G}_0, \sigma_{\text{Em}}^2, \lambda, \sigma_{\text{SY}}^2, \sigma_{\text{Eo}}^2 \mid \mathbf{y}, \mathbf{G}_0^*, s_{\text{Em}}^2, s_{\text{Eo}}^2, \mathbf{v}_g, \mathbf{v}_{\text{Em}}, s_{\text{sy}}^2, \mathbf{v}_{\text{sy}}, \mathbf{v}_{\text{Eo}}) \propto$$

$$p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma_{\text{Eo}}^2) \times p(\boldsymbol{\beta} \mid \mathbf{K}) \times p(\mathbf{a} \mid \mathbf{G}_0) \times p(\mathbf{G}_0 \mid \mathbf{v}_g, \mathbf{G}_0^*) \times p(\mathbf{e}_m \mid \lambda, \sigma_{\text{Em}}^2) \times \quad [2.14]$$

$$p(\sigma_{\text{Em}}^2 \mid \mathbf{v}_{\text{Em}}, s_{\text{Em}}^2) \times p(\lambda) \times p(\mathbf{h} \mid \sigma_{\text{SY}}^2) \times p(\sigma_{\text{SY}}^2 \mid \mathbf{v}_{\text{sy}}, s_{\text{sy}}^2) \times p(\sigma_{\text{Eo}}^2 \mid \mathbf{v}_{\text{Eo}}, s_{\text{Eo}}^2)$$

Sea $\boldsymbol{\Omega}$ el vector de componentes de varianza de los parámetros de posición de un modelo. Por ejemplo para GESY tenemos que:

$$\boldsymbol{\Omega}' = [\sigma_{\text{Ao}}^2, \sigma_{\text{AoAm}}^2, \sigma_{\text{Am}}^2, \sigma_{\text{AoAn}}^2, \sigma_{\text{An}}^2, \sigma_{\text{AmAn}}^2, \lambda, \sigma_{\text{Em}}^2, \sigma_{\text{SY}}^2, \sigma_{\text{Eo}}^2], \quad [2.15]$$

Sea además $\boldsymbol{\theta}_{-i}$ el vector de parámetros de posición que incluye los mismos elementos que $\boldsymbol{\theta}_i$ con excepción del *i*-ésimo. Entonces, la distribución condicional *a posteriori* de los vectores $\boldsymbol{\beta}$, \mathbf{a} , \mathbf{e}_m y \mathbf{h} es:

$$\left. \begin{array}{l} \boldsymbol{\beta} \\ a_o \\ a_m \\ a_n \\ e_m \\ h \end{array} \right| \boldsymbol{\Omega}, \mathbf{y} \sim N_{p+l+q+d+s} \left(\left[\begin{array}{c} \hat{\boldsymbol{\beta}} \\ \hat{a}_o \\ \hat{a}_m \\ \hat{a}_n \\ \hat{e}_m \\ \hat{\mathbf{h}} \end{array} \right], \left[\begin{array}{cccccc} \mathbf{X}'\mathbf{X} + \mathbf{K}^{-1} & \mathbf{X}'\mathbf{Z}_o & \mathbf{X}'\mathbf{Z}_m & \mathbf{X}'\mathbf{Z}_n & \mathbf{X}'\mathbf{Z}_p & \mathbf{X}'\mathbf{Z}_{\text{sy}} \\ \mathbf{Z}_o'\mathbf{X} & \mathbf{Z}_o'\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{11} & \mathbf{Z}_o'\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{12} & \mathbf{Z}_o'\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{13} & \mathbf{Z}_o'\mathbf{Z}_p & \mathbf{Z}_o'\mathbf{Z}_{\text{sy}} \\ \mathbf{Z}_m'\mathbf{X} & \mathbf{Z}_m'\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{21} & \mathbf{Z}_m'\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{22} & \mathbf{Z}_m'\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{23} & \mathbf{Z}_m'\mathbf{Z}_p & \mathbf{Z}_m'\mathbf{Z}_{\text{sy}} \\ \mathbf{Z}_n'\mathbf{X} & \mathbf{Z}_n'\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{31} & \mathbf{Z}_n'\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{32} & \mathbf{Z}_n'\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{33} & \mathbf{Z}_n'\mathbf{Z}_p & \mathbf{Z}_n'\mathbf{Z}_{\text{sy}} \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_o & \mathbf{Z}_p'\mathbf{Z}_m & \mathbf{Z}_p'\mathbf{Z}_n & \mathbf{Z}_p'\mathbf{Z}_p + \mathbf{E}_d^{-1} \frac{\sigma_{\text{Eo}}^2}{\sigma_{\text{Em}}^2} & \mathbf{Z}_p'\mathbf{Z}_{\text{sy}} \\ \mathbf{Z}_{\text{sy}}'\mathbf{X} & \mathbf{Z}_{\text{sy}}'\mathbf{Z}_o & \mathbf{Z}_{\text{sy}}'\mathbf{Z}_m & \mathbf{Z}_{\text{sy}}'\mathbf{Z}_n & \mathbf{Z}_{\text{sy}}'\mathbf{Z}_p & \mathbf{Z}_{\text{sy}}'\mathbf{Z}_{\text{sy}} + \mathbf{I}_s^{-1} \frac{\sigma_{\text{Eo}}^2}{\sigma_{\text{SY}}^2} \end{array} \right]^{-1} \right) \quad [2.16]$$

donde $\begin{bmatrix} \mathbf{g}^{11} & \mathbf{g}^{12} & \mathbf{g}^{13} \\ \mathbf{g}^{21} & \mathbf{g}^{22} & \mathbf{g}^{23} \\ \mathbf{g}^{31} & \mathbf{g}^{32} & \mathbf{g}^{33} \end{bmatrix} = \begin{bmatrix} \sigma_{Ao}^2 & \sigma_{AoAm} & \sigma_{AoAn} \\ \sigma_{AoAm} & \sigma_{Am}^2 & \sigma_{AmAn} \\ \sigma_{AoAn} & \sigma_{AmAn} & \sigma_{An}^2 \end{bmatrix}^{-1} \sigma_{Eo}^2$ mientras que $\hat{\boldsymbol{\beta}}, \hat{\mathbf{a}}_o, \hat{\mathbf{a}}_m, \hat{\mathbf{a}}_n, \hat{\mathbf{e}}_m$ y

$\hat{\mathbf{h}}$ son las soluciones del sistema de ecuaciones de modelo mixto representado en [2.17]:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} + \mathbf{K}^{-1} & \mathbf{X}'\mathbf{Z}_o & \mathbf{X}'\mathbf{Z}_m & \mathbf{X}'\mathbf{Z}_n & \mathbf{X}'\mathbf{Z}_p & \mathbf{X}'\mathbf{Z}_{sy} \\ \mathbf{Z}_o'\mathbf{X} & \mathbf{Z}_o'\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{11} & \mathbf{Z}_o'\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{12} & \mathbf{Z}_o'\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{13} & \mathbf{Z}_o'\mathbf{Z}_p & \mathbf{Z}_o'\mathbf{Z}_{sy} \\ \mathbf{Z}_m'\mathbf{X} & \mathbf{Z}_m'\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{21} & \mathbf{Z}_m'\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{22} & \mathbf{Z}_m'\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{23} & \mathbf{Z}_m'\mathbf{Z}_p & \mathbf{Z}_m'\mathbf{Z}_{sy} \\ \mathbf{Z}_n'\mathbf{X} & \mathbf{Z}_n'\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{31} & \mathbf{Z}_n'\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{32} & \mathbf{Z}_n'\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{33} & \mathbf{Z}_n'\mathbf{Z}_p & \mathbf{Z}_n'\mathbf{Z}_{sy} \\ \mathbf{Z}_p'\mathbf{X} & \mathbf{Z}_p'\mathbf{Z}_o & \mathbf{Z}_p'\mathbf{Z}_m & \mathbf{Z}_p'\mathbf{Z}_n & \mathbf{Z}_p'\mathbf{Z}_p + \mathbf{E}_d^{-1} \frac{\sigma_{Eo}^2}{\sigma_{Em}^2} & \mathbf{Z}_p'\mathbf{Z}_{sy} \\ \mathbf{Z}_{sy}'\mathbf{X} & \mathbf{Z}_{sy}'\mathbf{Z}_o & \mathbf{Z}_{sy}'\mathbf{Z}_m & \mathbf{Z}_{sy}'\mathbf{Z}_n & \mathbf{Z}_{sy}'\mathbf{Z}_p & \mathbf{Z}_{sy}'\mathbf{Z}_{sy} + \mathbf{I}_s^{-1} \frac{\sigma_{Eo}^2}{\sigma_{SY}^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}}_o \\ \hat{\mathbf{a}}_m \\ \hat{\mathbf{a}}_n \\ \hat{\mathbf{e}}_m \\ \hat{\mathbf{h}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}_o'\mathbf{y} \\ \mathbf{Z}_m'\mathbf{y} \\ \mathbf{Z}_n'\mathbf{y} \\ \mathbf{Z}_p'\mathbf{y} \\ \mathbf{Z}_{sy}'\mathbf{y} \end{bmatrix} \quad [2.17]$$

Asimismo, las distribuciones condicionales *a posteriori* de los componentes de varianza σ_{Em}^2 , σ_{SY}^2 y σ_{Eo}^2 son:

$$p(\sigma_{Em}^2 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y}) \propto (\sigma_{Em}^2)^{-\left(\frac{d + v_{Em}}{2} + 1\right)} \exp\left[-\frac{\mathbf{e}_m' \mathbf{E}^{-1} \mathbf{e}_m + v_{Em} s_{Em}^2}{2 \sigma_{Em}^2}\right] \quad [2.18]$$

$$p(\sigma_{SY}^2 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y}) \propto (\sigma_{SY}^2)^{-\left(\frac{s + v_{sy}}{2} + 1\right)} \exp\left[-\frac{\mathbf{h}'\mathbf{h} + v_{sy} s_{sy}^2}{2 \sigma_{SY}^2}\right] \quad [2.19]$$

$$p(\sigma_{Eo}^2 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y}) \propto (\sigma_{Eo}^2)^{-\left(\frac{\tilde{v}_{Eo}}{2} + 1\right)} \exp\left[-\frac{\tilde{v}_{Eo} \tilde{s}_{Eo}^2}{2 \sigma_{Eo}^2}\right] \quad [2.20]$$

Las ecuaciones [2.18], [2.19] y [2.20] corresponden a distribuciones χ^2 escaladas con grados de credibilidad: $\tilde{v}_{Em} = d + v_{Em}$; $\tilde{v}_{sy} = s + v_{sy}$; $\tilde{v}_{Eo} = v_{Eo} + n$; y parámetros de escala iguales a

$$\begin{aligned} \tilde{s}_{Em}^2 &= \frac{\mathbf{e}_m' \mathbf{E}^{-1} \mathbf{e}_m + v_{Em} s_{Em}^2}{v_{Em} + d}; \\ \tilde{s}_{sy}^2 &= \frac{\mathbf{h}'\mathbf{h} + v_{sy} s_{sy}^2}{v_{sy} + s}, \\ \tilde{s}_{Eo}^2 &= \frac{(\mathbf{y} - \mathbf{W}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{W}\boldsymbol{\theta}) + v_{Eo} s_{Eo}^2}{\tilde{v}_{Eo}} \end{aligned} \quad [2.21]$$

En forma más compacta, las citadas distribuciones son iguales a

$$\begin{aligned} \sigma_{Em}^2 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y} &\sim \chi^{-2}(\tilde{v}_{Em}, \tilde{s}_{Em}^2) \\ \sigma_{SY}^2 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y} &\sim \chi^{-2}(\tilde{v}_{sy}, \tilde{s}_{sy}^2) \\ \sigma_{Eo}^2 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y} &\sim \chi^{-2}(\tilde{v}_{Eo}, \tilde{s}_{Eo}^2) \end{aligned} \quad [2.22]$$

2.2.3.3. Muestreo de \mathbf{G}_0 utilizando distintos grados de credibilidad

La estimación Bayesiana de matrices de (co)varianzas genéticas se suele realizar empleando una formulación conjugada, mediante la distribución Wishart invertida. Un problema con este enfoque es que los grados de credibilidad deben ser iguales para todos los elementos de la matriz. Esto es inconveniente cuando la información para ciertos parámetros es mayor que para otros. Un ejemplo de ello es el modelo con efectos aditivos de abuela materna, donde para estimar la varianza aditiva por efectos directos la base de datos suele contener mucha más información que para los efectos maternos. A su vez, estos últimos poseen más información que los efectos de abuela materna (Cantet y Suárez, 2009; Munilla y Cantet, 2012). Para desarrollar el muestreo de la densidad condicional posterior de \mathbf{G}_0 , definiremos \mathbf{S}_g como

$$\mathbf{S}_g = \mathbf{a}'(\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1})\mathbf{a} = \begin{bmatrix} \mathbf{a}_o' \mathbf{A}^{-1} \mathbf{a}_o & \mathbf{a}_o' \mathbf{A}^{-1} \mathbf{a}_m & \mathbf{a}_o' \mathbf{A}^{-1} \mathbf{a}_n \\ \mathbf{a}_m' \mathbf{A}^{-1} \mathbf{a}_o & \mathbf{a}_m' \mathbf{A}^{-1} \mathbf{a}_m & \mathbf{a}_m' \mathbf{A}^{-1} \mathbf{a}_n \\ \mathbf{a}_n' \mathbf{A}^{-1} \mathbf{a}_o & \mathbf{a}_n' \mathbf{A}^{-1} \mathbf{a}_m & \mathbf{a}_n' \mathbf{A}^{-1} \mathbf{a}_n \end{bmatrix} \quad [2.23]$$

De este modo, podemos especificar que la matriz de covarianzas aditivas es proporcional a la distribución siguiente:

$$p(\mathbf{G}_0 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y}) \propto |\mathbf{G}_0|^{-\frac{1}{2}(v+q+l+1)} \exp\left\{-\frac{1}{2}\text{tr}[\mathbf{G}_0^{-1}(\mathbf{S}_g + \mathbf{G}_0^{*-1})]\right\} \quad [2.24]$$

La expresión [2.24] es el kernel de una distribución de orden $l \times l$, Wishart invertida escalada con grados de credibilidad igual a $v + q$ y matriz de escala $\mathbf{S}_g + \mathbf{G}_0^{*-1}$, que es equivalente a escribir:

$$\mathbf{G}_0 | \boldsymbol{\theta}, \boldsymbol{\Omega}_{-i}, \mathbf{y} \sim \text{IW}_l(v+q, (\mathbf{S}_g + \mathbf{G}_0^{*-1})^{-1}) \quad [2.25]$$

Para reflejar las diferencias en la incertidumbre de las varianzas y covarianzas de los efectos genéticos aditivos, los elementos de la matriz \mathbf{G}_0 fueron muestreados individualmente usando una distribución Wishart Invertida Generalizada (GIW, Brown, 2001), según lo sugerido por Daniels y Pourahmadi (2002). La distribución GIW se obtiene a partir de la descomposición de Bartlett de una matriz simétrica y positiva-definida, y tiene un número mayor de hiperparámetros que la Wishart invertida. Sea la matriz de varianzas y covarianzas genéticas igual a:

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_{11} & \mathbf{g}_{12} & \mathbf{g}_{13} \\ \mathbf{g}_{21} & \mathbf{g}_{22} & \mathbf{g}_{23} \\ \mathbf{g}_{31} & \mathbf{g}_{32} & \mathbf{g}_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_{11} & \mathbf{g}_1 \\ \mathbf{g}_1' & \mathbf{G}_{-1} \end{bmatrix} \quad [2.26]$$

donde 1 indica efectos directos, 2 a los maternos y 3 a los de abuela materna. Aquí, \mathbf{g}_1 es un vector de orden 2×1 y \mathbf{G}_{-1} una matriz de 2×2 . Entonces, la descomposición de Bartlett de \mathbf{G} permite escribir $\mathbf{G} = \mathbf{T} \boldsymbol{\Delta} \mathbf{T}'$ donde

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{g}_{11} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} \mathbf{g}_{22} & \mathbf{g}_{23} \\ \mathbf{g}_{32} & \mathbf{g}_{33} \end{bmatrix} - \begin{bmatrix} \mathbf{g}_{12} \\ \mathbf{g}_{13} \end{bmatrix} (\mathbf{g}_{11})^{-1} \begin{bmatrix} \mathbf{g}_{12} & \mathbf{g}_{13} \end{bmatrix}' \end{bmatrix} \quad \mathbf{T} = \begin{bmatrix} 1 & \mathbf{0} \\ \begin{bmatrix} (\mathbf{g}_{11})^{-1} \mathbf{g}_{12} \\ (\mathbf{g}_{11})^{-1} \mathbf{g}_{13} \end{bmatrix} & \mathbf{I}_2 \end{bmatrix} \quad [2.27]$$

Mediante la descomposición de Bartlett y haciendo uso de las propiedades de la GIW, g_{11} y $\begin{bmatrix} g_{22} & g_{23} \\ g_{32} & g_{33} \end{bmatrix} - \begin{bmatrix} g_{12} \\ g_{13} \end{bmatrix} (g_{11})^{-1} [g_{12} \ g_{13}]'$ se distribuyen siguiendo una χ^2 mientras que, $\begin{bmatrix} (g_{11})^{-1} g_{12} \\ (g_{11})^{-1} g_{13} \end{bmatrix}$ sigue una distribución normal bivariada. En primer lugar se debe muestrear g_{11} , luego se procede a muestrear g_{12} condicional al valor muestreado en g_{11} . Luego se muestrea g_{22} condicional a los valores muestreados de g_{11} y g_{12} , y así sucesivamente. Esta formulación permite usar distintos grados de credibilidad para las tres varianzas aditivas (σ_{Ao}^2 , σ_{Am}^2 , σ_{An}^2). En la presente tesis, los grados de credibilidad se fijaron de modo tal que σ_{Ao}^2 recibió dos veces más credibilidad que σ_{Am}^2 y seis veces más credibilidad que σ_{An}^2 ($v_{Ao} = 30$, $v_{Am} = 15$ y $v_{An} = 5$).

2.2.3.4. Muestreo de λ

Para el parámetro λ , Quintanilla et al. (1999) propusieron utilizar una distribución *a priori* uniforme: $\lambda \sim U(-1, 1)$. Sin embargo, al obtener la densidad condicional posterior de este parámetro, hallaron que no fue proporcional a ninguna distribución conocida. Dado que el muestreo de Gibbs requiere las expresiones exactas de las densidades condicionales posteriores de los parámetros a estimar, de las que se muestrea en forma iterativa hasta obtener las distribuciones marginales, los autores recurrieron a un algoritmo de Metrópolis-Hastings por etapas dentro del muestreo de Gibbs, descrito en el libro de Tanner (1996). En contraste, en la presente investigación el muestreo se realizó a partir de una distribución normal truncada. Si bien Chib (1993) es la primera referencia en el tema, Heringstad et al. (2003) fueron los primeros en aplicar este tipo de muestreo en un modelo de umbral (*threshold model*) para datos de mastitis clínica. En esencia, el método se basa en particionar el vector e_m en $e_m = [e_{mD}', e_{mP}']'$ de modo semejante al vector de valores de cría del modelo animal reducido (Quaas y Pollak, 1980), en donde los individuos se los considera o padres o no-padres (*parents or non-parents*). Además, al igual que en el modelo reducido, es condición necesaria que los sub-vectores resultantes se ordenen de modo que las madres de las vacas con progenie (agrupadas en e_{mD}) antecedan a las vacas que no sean madres de vacas (agrupadas en e_{mP}). Por lo tanto, la densidad considerada fue: $p(e_{mD}, e_{mP} | E_\lambda) \sim N_d(\theta, E_\lambda \sigma_{Em}^2)$. En orden de proceder con el algoritmo, el primer paso consiste en ajustar la siguiente regresión:

$$e_{mP} = \lambda S e_{mD} + \epsilon \quad [2.28]$$

donde la matriz S relaciona los efectos ambientales maternos permanentes (e_m) de una vaca con los de su madre. El error de regresión (ϵ) sigue $\epsilon \sim N_d(\theta, I \sigma_\epsilon^2)$ con $\sigma_\epsilon^2 = (1 - \lambda) \sigma_{Em}^2$. La varianza del error de regresión σ_ϵ^2 , por su parte, fue muestreada de la siguiente densidad $p(\sigma_\epsilon^2 | e_m, \lambda) \sim \epsilon' \epsilon \chi_{d-2}^{-2}$ (Heringstad et al., 2003). Finalmente, el parámetro λ se muestreó de la siguiente densidad condicional posterior truncada

$$p(\lambda | e_{mD}, e_{mP}, \sigma_\epsilon^2) \sim N_{TR}(E(\lambda), \text{Var}(\lambda)) \quad [2.29]$$

con parámetros

$$E(\lambda) = (\mathbf{e}_{mD}' \mathbf{S}' \mathbf{S} \mathbf{e}_{mD})^{-1} \mathbf{e}_{mD}' \mathbf{S}' \mathbf{e}_{mP} \quad \text{Var}(\lambda) = (\mathbf{e}_{mD}' \mathbf{S}' \mathbf{S} \mathbf{e}_{mD})^{-1} \sigma_{\epsilon}^2 \quad [2.30]$$

Para garantizar que la matriz \mathbf{E} resultante sea positiva definida, se restringió el espacio paramétrico de λ . Se encontró, por prueba y error, que cuando $-0.30 < \lambda < -0.01$ para los modelos E y ESY, $-0.25 < \lambda < -0.01$ para los modelos GE y GESY, la matriz \mathbf{E} es positiva definida, es decir $\mathbf{E} > 0$. El algoritmo de aceptación-rechazo que se utilizó para tomar las muestras de la distribución normal doble truncada fue el propuesto por Robert (1995).

2.2.3.5. Esquema de muestreo para el algoritmo de GS en el modelo mixto

El algoritmo GS para el modelo [2.1] se desarrolla de la siguiente manera:

- I. Se asignan valores arbitrarios para β , \mathbf{a} , \mathbf{e}_m , \mathbf{h} , σ_{Ao}^2 , σ_{Am}^2 , σ_{An}^2 , σ_{AoAm} , σ_{AoAn} , σ_{AmAn} , σ_{Em}^2 , λ , σ_{SY}^2 y σ_{Eo}^2 .
- II. Se muestrean β , \mathbf{a} , \mathbf{e}_m y \mathbf{h} a partir de $p(\theta_i | \theta_i, \Omega, \mathbf{y})$ en [2.16].
- III. Se muestrea σ_{Eo}^2 a partir de $p(\sigma_{Eo}^2 | \theta, \Omega_{-i}, \mathbf{y})$ en [2.20].
- IV. Se muestrea σ_{Em}^2 a partir de $p(\sigma_{Em}^2 | \theta, \Omega_{-i}, \mathbf{y})$ ([2.18]).
- V. Se muestrea λ a partir de $p(\lambda | \mathbf{e}_{mD}, \mathbf{e}_{mP}, \sigma_{\epsilon}^2)$ en [2.29], siguiendo a Chib (1993).
- VI. Se muestrea \mathbf{G}_o a partir de una distribución GIW $p(\mathbf{G}_o | \theta, \Omega_{-i}, \mathbf{y})$ en [2.25], haciendo uso de la descomposición de Bartlett.
- VII. Se muestrea σ_{SY}^2 a partir de $p(\sigma_{SY}^2 | \theta, \Omega_{-i}, \mathbf{y})$ en [2.19].

Se repiten los pasos II a VII, en número igual al largo de la cadena a generar.

Jensen et al. (1994) produjeron una descripción más detallada de la implementación del muestreo de Gibbs en modelos animales con efectos maternos y efectos ambientales maternos permanentes que son independientes entre sí.

2.2.3.6. Muestreo de Gibbs

El muestreo de Gibbs fue implementado mediante una cadena larga de 100,000 ciclos. En consonancia con Geyer (1992), las primeras 2,000 iteraciones (2%) se descartaron debido al período de calentamiento (*burn-in*), y las restantes se utilizaron para la estimación de las densidades marginales posteriores. Los estadísticos posteriores tales como la media, la mediana, el modo, el desvío estándar y el intervalo de alta densidad posterior al 95% (IADP95) se obtuvieron con el programa POSTGIBBSF90, proveniente del paquete BLUPF90 (Misztal et al., 2002). La convergencia de todos los parámetros fue evaluada a través de la observación de los promedios acumulados por iteración, y la prueba del estadístico de Geweke (Geweke, 1992), a través del paquete de programas “*Bayesian Output Analysis*” (BOA) versión 1.1.7 (Smith, 2003, 2007). Todas las cadenas convergieron con el número de iteraciones establecidas. Sin embargo, para que el modelo ISY alcance convergencia, se extendió la cadena a 178000 ciclos. Dado

lo inédito del ajuste de modelos propuestos, los programas para las estimaciones fueron escritos en lenguaje FORTRAN. Por último, con el objetivo de evaluar si hubo un cambio en el ordenamiento de los animales evaluados, se calcularon los coeficientes por rangos de Spearman tanto para los valores de cría directos como para los maternos, en cada uno de los ocho modelos bajo estudio.

Con el fin de comparar los parámetros de dispersión estimados con las realizadas por otros autores, los resultados se expresaron en términos de heredabilidades y correlaciones en lugar de varianzas y correlaciones. De este modo, tenemos las siguientes parametrizaciones:

$$h_o^2 = \frac{\sigma_{Ao}^2}{\sigma^2}; \quad h_m^2 = \frac{\sigma_{Am}^2}{\sigma^2}; \quad h_n^2 = \frac{\sigma_{An}^2}{\sigma^2}; \quad r_{AoAm} = \frac{\sigma_{AoAm}}{\sqrt{\sigma_{Ao}^2 \sigma_{Am}^2}}; \quad [2.31]$$

$$r_{AoAn} = \frac{\sigma_{AoAn}}{\sqrt{\sigma_{Ao}^2 \sigma_{An}^2}}; \quad r_{AmAn} = \frac{\sigma_{AmAn}}{\sqrt{\sigma_{Am}^2 \sigma_{An}^2}}; \quad c^2 = \frac{\sigma_{Em}^2}{\sigma^2}; \quad c_{SY} = \frac{\sigma_{SY}^2}{\sigma^2};$$

Por otro lado, los estimadores de heredabilidad total (h_T^2) se calcularon a partir de las estimaciones de los componentes de (co)varianza usando la siguiente fórmula $h_T^2 = \frac{\sigma_T^2}{\sigma^2}$ donde σ^2 es la varianza fenotípica del modelo y, σ_T^2 es la varianza aditiva total (Dickerson, 1947). En este caso, σ_T^2 fue calculada de acuerdo a Willham (1972), quien establece que para modelos con efectos maternos $\sigma_T^2 = \sigma_{Ao}^2 + 1,5 \sigma_{AoAm} + 0,5 \sigma_{Am}^2$.

2.3. RESULTADOS Y DISCUSIÓN

El esquema de muestreo presentó bajas autocorrelaciones para todas las variables y el algoritmo de Gibbs convergió rápido en todos los casos, con excepción de la variable c_{SY} del modelo ISY que requirió un número mayor de iteraciones. En el Cuadro 2.2 se muestran las medias marginales posteriores con sus respectivos desvíos estándares y los IADP95 de los parámetros de dispersión genéticos y ambientales de los ocho modelos.

Cuadro 2.2. Medias marginales posteriores con sus respectivos desvíos estándares, e intervalos de alta densidad posterior (95% IADP, en paréntesis) para los parámetros de dispersión estimados en los ocho modelos propuestos.

Parámetros											
Modelos	h_o^2	h_m^2	h_n^2	r_{AoAm}	r_{AoAn}	r_{AmAn}	λ	c^2	c_{SY}	σ_{Eo}^2	h_T^2
I	0.16 ± 0.02	0.09 ± 0.02	*	-0.71 ± 0.08	*	*	*	0.15 ± 0.02	*	495.99 ± 7.34	0.077
	(0.12 - 0.20)	(0.05 - 0.14)		(-0.86 - -0.52)				(0.12 - 0.18)		(481.75 - 510.49)	
E	0.19 ± 0.00	0.15 ± 0.00	*	-0.21 ± 0.08	*	*	-0.28 ± 0.03	0.07 ± 0.01	*	487.31 ± 6.56	0.213
	(0.18 - 0.20)	(0.14 - 0.16)		(-0.37 - -0.05)			(-0.30 - -0.20)	(0.05 - 0.09)		(474.57 - 500.32)	
G	0.16 ± 0.00	0.13 ± 0.00	0.11 ± 0.00	-0.21 ± 0.00	0.09 ± 0.10	0.19 ± 0.10	*	0.07 ± 0.00	*	482.09 ± 6.45	0.181
	(0.15 - 0.17)	(0.12 - 0.14)	(0.10 - 0.11)	(-0.37 - -0.05)	(-0.10 - 0.29)	(-7.10 ⁻³ - 0.40)		(0.06 - 0.09)		(469.68 - 494.95)	
GE	0.17 ± 0.00	0.13 ± 0.00	0.10 ± 0.00	-0.21 ± 0.08	0.10 ± 0.10	0.20 ± 0.11	-0.21 ± 0.04	0.05 ± 0.00	*	483.88 ± 6.49	0.187
	(0.16 - 0.18)	(0.13 - 0.14)	(0.09 - 0.11)	(-0.37 - -0.05)	(-0.11 - 0.30)	(-6.10 ⁻³ - 0.40)	(-0.25 - -0.08)	(0.03 - 0.06)		(471.36 - 496.76)	
ISY	0.11 ± 0.02	0.08 ± 0.02	*	-0.63 ± 0.13	*	*	*	0.15 ± 0.02	0.03 ± 0.01	497.64 ± 7.41	0.062
	(0.06 - 0.17)	(0.05 - 0.13)		(-0.85 - -0.33)				(0.11 - 0.18)	(0.01 - 0.05)	(483.15 - 512.23)	
ESY	0.19 ± 0.00	0.15 ± 0.00	*	-0.21 ± 0.08	*	*	-0.28 ± 0.03	0.07 ± 0.01	0.01 ± 0.00	484.68 ± 6.64	0.210
	(0.18 - 0.20)	(0.14 - 0.16)		(-0.37 - -0.05)			(-0.30 - -0.20)	(0.05 - 0.09)	(2.10 ⁻² - 0.03)	(471.77 - 497.82)	
GSY	0.16 ± 0.00	0.13 ± 0.00	0.10 ± 0.00	-0.21 ± 0.08	0.10 ± 0.10	0.20 ± 0.11	*	0.07 ± 0.00	0.01 ± 0.00	479.92 ± 6.53	0.181
	(0.15 - 0.17)	(0.12 - 0.14)	(0.09 - 0.11)	(-0.37 - -0.05)	(-0.11 - 0.30)	(-7.10 ⁻³ - 0.41)		(0.06 - 0.09)	(1.10 ⁻³ - 0.02)	(467.25 - 492.78)	
GESY	0.17 ± 0.00	0.13 ± 0.00	0.10 ± 0.00	-0.21 ± 0.08	0.10 ± 0.10	0.20 ± 0.10	-0.21 ± 0.04	0.05 ± 0.00	0.01 ± 0.00	481.69 ± 6.52	0.186
	(0.16 - 0.18)	(0.12 - 0.14)	(0.09 - 0.11)	(-0.37 - -0.05)	(-0.11 - 0.30)	(-0.11 - 0.41)	(-0.25 - -0.08)	(0.03 - 0.06)	(5.10 ⁻⁴ - 0.02)	(469.11 - 494.60)	

h_o^2 = heredabilidad directa; h_m^2 = heredabilidad materna; h_n^2 = heredabilidad de abuela materna; r_{AoAm} = correlación genética directa - materna; r_{AoAn} = correlación genética directa - de abuela materna; r_{AmAn} = correlación genética materna - de abuela materna; λ = correlación entre los efectos ambientales maternos permanentes de una madre y sus hijas; c^2 = fracción de la varianza fenotípica debida a los efectos ambientales maternos permanentes; c_{SY} = proporción de la varianza fenotípica que es debida a la interacción toro \times año; σ_{Eo}^2 = varianza residual; h_T^2 = heredabilidad total; * = parámetros que no pertenecen al modelo referido.

Las densidades marginales posteriores fueron unimodales y tendieron a ser simétricas. Sólo los parámetros λ y c_{SY} presentaron densidades marginales posteriores asimétricas (véase Figuras 2.1 y 2.2). Sin embargo, dado que en todos casos las medias, las medianas y los modos marginales posteriores fueron similares, en el Cuadro 2.2 solo se informan las medias posteriores. Los IADP95 de los parámetros h_o^2 , h_m^2 , h_n^2 , r_{AoAm} , r_{AmAn} , λ , c^2 , c_{SY} y σ_{Eo}^2 no incluyen el cero. Estos resultados sugieren que ninguno de esos parámetros es nulo. En los ocho modelos y consistentemente las estimaciones de h_o^2 fueron más grandes que h_m^2 y estas últimas, a su vez, mayores que h_n^2 . Las estimaciones obtenidas estuvieron en el rango de 0.11 a 0.19 para h_o^2 , 0.08 a 0.15 para h_m^2 , y 0.10 a 0.11 para h_n^2 . En un subarchivo de 935 registros de esta misma base de datos, Cantet et al. (1993) estimaron $h_o^2 = 0.12$ y $h_m^2 = 0.03$, mediante el método de estimación de Máxima Verosimilitud Restringida (REML; Patterson y Thompson, 1971). Adicionalmente, Dodenhoff et al. (1999a) encontraron valores de $h_o^2 = 0.22$, $h_m^2 = 0.11$, y $h_n^2 = 0.06$ en un rodeo Angus empleando REML. En la presente investigación, las estimaciones que presentaron mayor magnitud para h_o^2 y h_m^2 fueron observadas en los modelos E y ESY, mientras que en ISY se encontraron las estimaciones para ambos parámetros de menor magnitud (ver discusión de r_{AoAm} más adelante). Las estimaciones de h_m^2 de los modelos E, G y GE son mayores a las estimadas bajo el modelo materno I, mientras que lo opuesto ocurre con el parámetro c^2 : el valor más alto se encontró en los modelos I e ISY. Estos resultados son coincidentes con los trabajos de Meyer (1992), Robinson (1996), Dodenhoff et al. (1998), Quintanilla et al. (1999) e Iwaisaki et al. (2005), y sugieren que bajo el modelo materno convencional (modelo I) la h_m^2 está subestimada y c^2 sobreestimado. Finalmente, se debe indicar que los valores estimados de σ_{Eo}^2 de los ocho modelos fueron similares.

Figura 2.1. Distribuciones marginales posteriores de la correlación entre los efectos ambientales maternos permanentes madre-hija (λ) de los modelos E, ESY, GE y GESY.

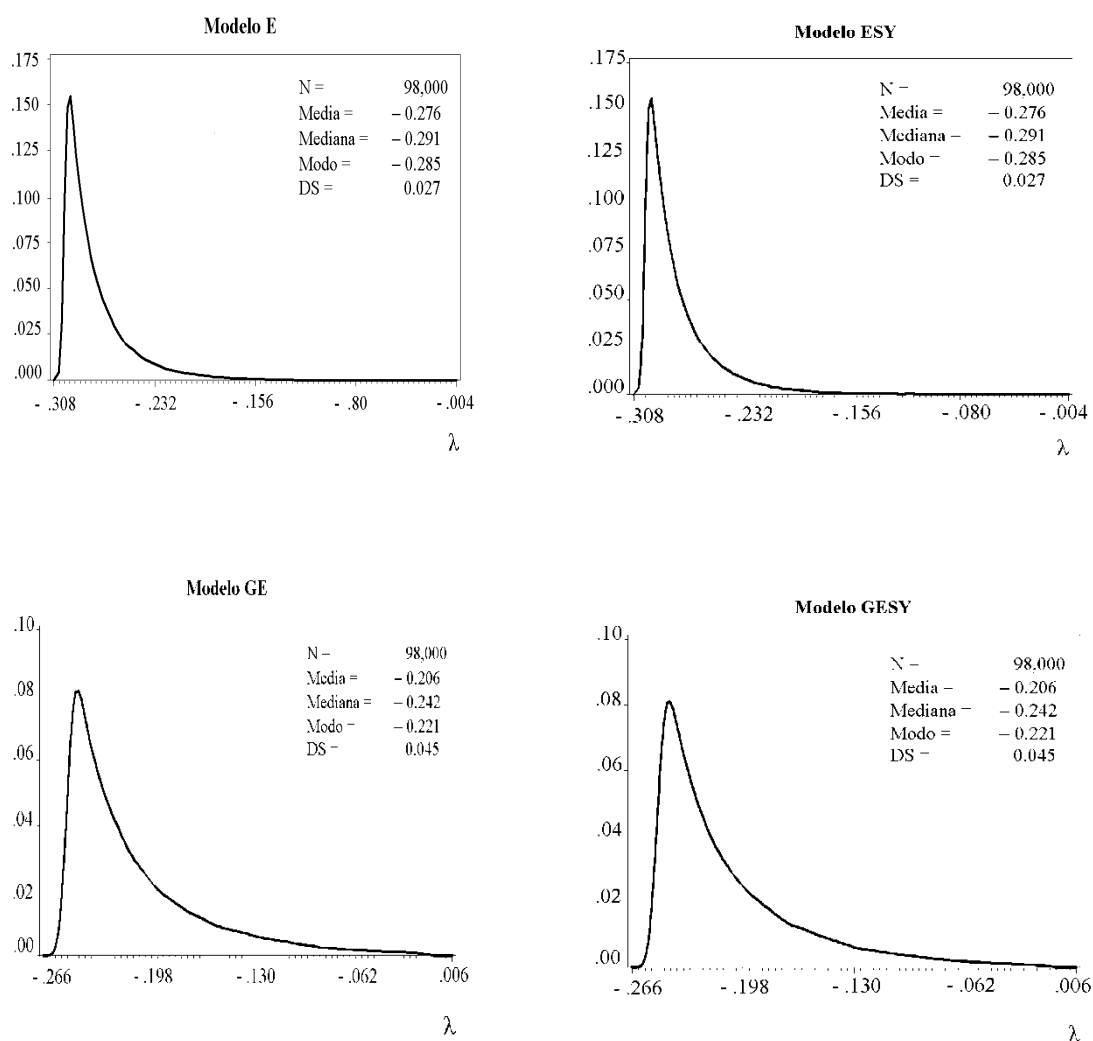
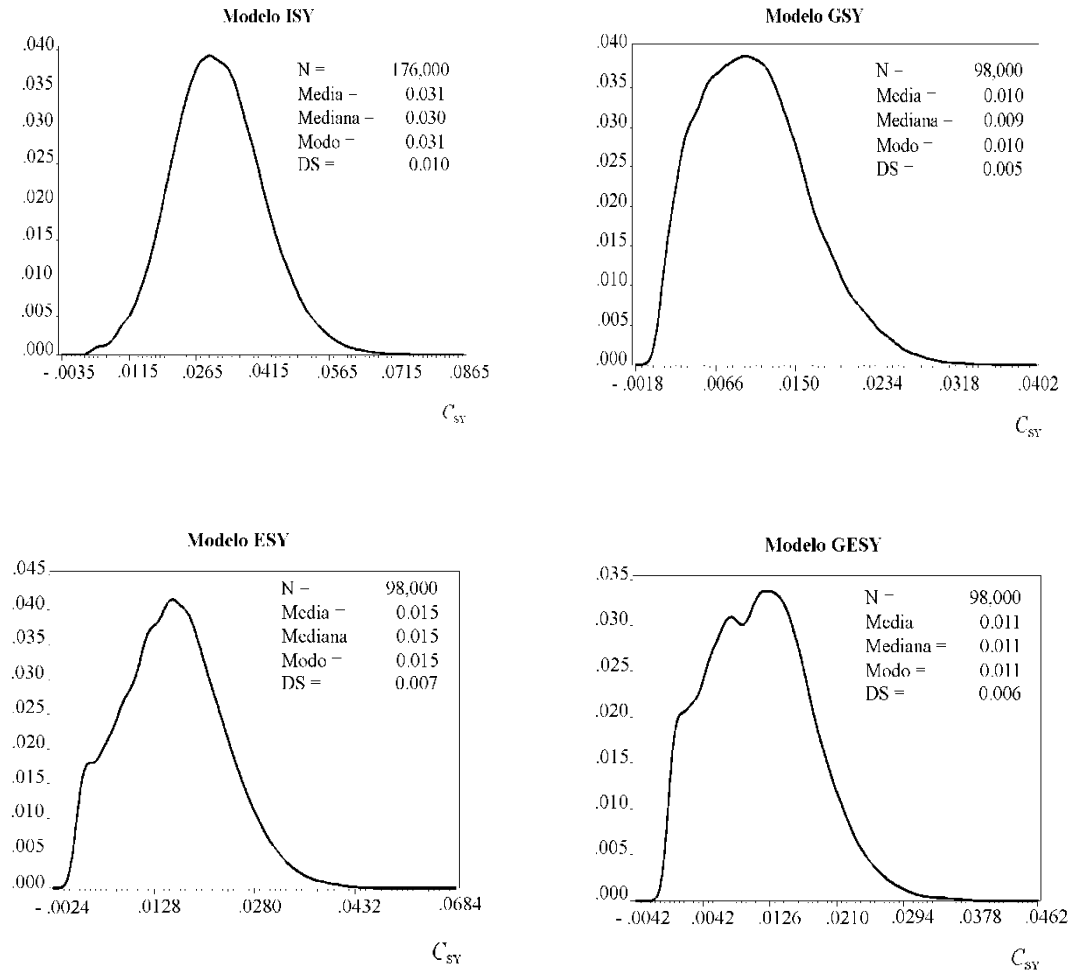


Figura 2.2. Distribuciones marginales posteriores de la proporción de la variancia fenotípica que es debida a la interacción toro \times año (c_{SY}) en los modelos ISY, ESY, GSY y GESY.



Como era de esperar, los signos de las estimaciones de r_{AoAm} fueron consistentemente negativos para todos los modelos ajustados. En aquellos modelos que consideran los efectos aditivos de abuela materna y/o se asumen que los efectos ambientales maternos permanentes no son independientes (E, G, GE, ESY, y GESY), las magnitudes de las estimaciones de r_{AoAm} fueron tres veces menores que en el modelo materno convencional I (-0.21 vs. -0.71). La inclusión de la interacción toro \times año en ISY redujo la magnitud negativa de r_{AoAm} en 0.08 unidades, pero disminuyó también la heredabilidad directa (h_o^2) y la materna (h_m^2). En cambio, los modelos que incluyeron la correlación entre los efectos ambientales maternos permanentes de una madre y sus hijas, o los efectos aditivos de abuela materna (modelos E y G), no solo redujeron la magnitud negativa de r_{AoAm} en 0.50 unidades, sino que además produjeron estimaciones más favorables de h_o^2 y h_m^2 .

2.3.1. Efectos aditivos de abuela materna

Los valores estimados de h_n^2 sugieren que los efectos aditivos de abuela materna existen, y son importantes, para el carácter peso al destete de este rodeo Angus. Bajo el modelo G, los valores de h_m^2 y r_{AoAm} fueron 0.04 y 0.50 unidades superiores, mientras que h_o^2 resultó igual, respecto las estimaciones obtenidas con el modelo I. Por su parte, Dodenhoff et al. (1998, 1999a,b) intentaron estimar los nueve componentes de (co)varianzas en el modelo de abuela materna presentado por Willham (1972) el cual incluye los efectos aditivos y ambientales permanentes de abuela materna (como primera generalización a otro tipo de antecesores), en distintas bases de datos provenientes de razas puras y compuestas mediante REML usando un algoritmo de *Average Information*. Dada la complejidad del modelo, la convergencia sólo se alcanzó en ocho de las doce bases de datos analizadas. Con excepción de una línea selecta de la raza Hereford para el carácter peso al año (Dodenhoff et al., 1998), las estimaciones de r_{AoAm} para peso al destete en los modelos de abuela materna fueron en todos los casos, menos negativas que aquellas estimaciones obtenidas bajo el modelo materno convencional. Contrariamente a lo esperado por Willham (1972), las estimaciones de r_{AmAn} de la presente tesis fueron positivas en los cuatro casos estudiados. Sin embargo, los resultados aquí obtenidos fueron consistentes en todos los modelos con efectos aditivos de abuela materna (ver Cuadro 2.2, modelos GE, GSY y GESY). Finalmente, los IADP95 del parámetro r_{AoAm} incluyen al cero, indicando que en este archivo de datos su efecto es poco apreciable. Particularmente, este resultado coincide con Willham (1972) quién expresó sus dudas con respecto a la existencia de correlación entre los efectos directos y los de abuela materna. La heredabilidad total (h_T^2) en el modelo G fue mayor a la del I debido a la magnitud menos negativa de la covarianza entre los efectos directos y los maternos (σ_{AoAm}). La misma situación se presentó en el modelo E. A diferencia de Dodenhoff et al. (1998, 1999a,b), al ajustarse los efectos genéticos de abuela materna no se encontraron problemas con la convergencia de los parámetros. Es factible que la mejora en la performance de las cadenas de la presente investigación no haya sido debida al uso de otro método de estimación, sino a la estructura de los datos: el número de madres y abuelas maternas con registro propio, el número de crías por madre, el número de generaciones en las que se registraron los datos y las relaciones de parentesco de los animales, fundamentalmente por vía materna (Gerstmayr, 1992; Maniatis y Pollot, 2003). En la presente tesis se ignoraron los efectos ambientales permanentes de abuela materna para evitar la excesiva sobreparametrización de los modelos

2.3.2. Modelos con efectos ambientales

Al ajustar el modelo ISY, las estimaciones de h_o^2 y h_m^2 fueron menores y r_{AoAm} fue 0.08 unidades mayor (menos negativa), comparados con las estimaciones de los mismos parámetros obtenidas con el modelo I, en coincidencia con la investigación de Dodenhoff et al. (1999b). Es posible que esta reducción en h_o^2 y h_m^2 se deba al hecho que la interacción toro \times año captura parte de la variabilidad asociada a σ_{Ao}^2 y σ_{Am}^2 . Así, la correlación entre los valores muestreados de σ_{Ao}^2 y σ_{SY}^2 fue la de mayor magnitud y de signo negativo (-0.58 en ISY). El confundimiento entre σ_{Ao}^2 y σ_{SY}^2 fue más evidente

al intentar ajustarse (sin éxito) la interacción toro \times año con una estructura de covarianzas que consideraba las relaciones aditivas entre toros. Bajo estas condiciones, σ_{Ao}^2 (y por ende h_o^2) tendió a cero. Como era de esperar, las estimaciones de c^2 no estuvieron afectadas por la inclusión de la interacción toro \times año. Al incluir un efecto aleatorio de toro \times cabaña en animales Angus en Suiza, Baschnagel et al. (1999) obtuvieron resultados similares a los aquí observados. La variancia de la interacción toro \times año representó entre 1 y 3 % de la varianza fenotípica en esta base de datos. Por su parte, Lee y Pollak (1997) reportaron $c_{SY} = 0.03$ en terneros Simmental, y Dodenhoff et al. (1999b) estimaron valores de c_{SY} variables entre 0.02 a 0.10 en razas diversas. La estimación de h_T^2 proveniente del modelo ISY fue la de menor magnitud, sugiriendo que no es necesario ajustar al modelo la interacción toro \times año; particularmente cuando los efectos aditivos de abuela materna (G) o los efectos ambientales maternos permanentes correlacionados (E), ya se encuentran dentro del modelo. En este punto debe destacarse que el presente análisis utiliza datos provenientes de un único rodeo. Por ello, resultó imposible ajustar la interacción toro \times rodeo y, posteriormente, medir su impacto sobre la estimación de los parámetros. Si bien es factible que los efectos causantes de la interacción toro \times año sean distintos de aquellos que originan la interacción toro \times rodeo (por ejemplo, los efectos genéticos no aditivos), no necesariamente ambos tipos de interacciones redistribuyen la variancia genética de manera similar. Sin embargo, como la variancia entre toros explica un cuarto de σ_{Ao}^2 , es dable esperar que la inclusión de una interacción que involucre a los toros disminuya de algún modo la variancia aditiva.

El antagonismo ambiental entre los pesos al destete de vacas y los de sus hijas que luego se transforman en madres, se evidenció en estimaciones de λ negativas, coincidentemente con las otras dos estimaciones previas: -0.05 (Cantet et al., 1988) y -0.20 (Quintanilla et al., 1999). En el presente estudio las medias posteriores de λ fueron -0.28 ± 0.03 en los modelos E y ESY, y -0.21 ± 0.03 en los modelos GE y GESY. Aunque para los cuatro modelos involucrando a E las densidades posteriores de λ no fueron simétricas, las medianas y los modos marginales posteriores presentaron valores muy similares a las medias marginales posteriores. Las estimaciones de λ fueron consistentemente negativas y en el rango de -0.25 a -0.30 para todos los modelos (ver Figura 2.1). Para peso al destete, estos resultados pueden interpretarse como evidencias de una relación negativa entre los efectos ambientales maternos permanentes de dos generaciones consecutivas de madres. Por su parte, al analizar datos de la raza Bruna dels Pirineus, Quintanilla et al. (1999) estimaron λ en -0.19 ± 0.28 (media de la distribución posterior y desvío estándar; moda: -0.15). También encontraron que al asumir independencia entre los efectos ambientales maternos permanentes de dos generaciones consecutivas se producen estimaciones sesgadas de los componentes de (co)varianza. Bajo el mismo modelo, Iwaisaki et al. (2005) estimaron los parámetros de dispersión mediante un algoritmo REML “modificado” en el que la verosimilitud era calculada para una serie de valores determinados de λ ($\lambda = 0, -0.10, -0.20, -0.30$). Se halló que, a medida que el valor de λ se hacía más negativo, r_{AoAm} se reducía de -0.25 a -0.17 en Gelbvieh, y de -0.59 a -0.52 en Limousin. Además, encontraron en ambas razas que la función de verosimilitud restringida era maximizada con un valor del parámetro de -0.20 .

2.3.3. Modelos que combinan efectos genéticos de abuela materna y efectos ambientales

Al comparar de a pares modelos que diferían únicamente en la presencia de la interacción toro \times año (E y ESY, G y GSY, GE y GESY), se observaron medias marginales posteriores idénticas. Consecuentemente, la variabilidad asociada con la interacción toro \times año fue explicada por los efectos ancestrales ambientales maternos (E), por los efectos aditivos de abuela materna (G), o por ambos (GE). Cuando se ajustó la interacción toro \times año simultáneamente con los efectos ambientales maternos permanentes (modelo ESY) y los valores de cría de abuela materna (modelo GESY), las medias marginales posteriores de c_{SY} presentaron formas irregulares comparadas con la densidad marginal posterior de c_{SY} en ISY (Figura 2.2). Esto sugiere que la estructura de datos no posee información suficiente para estimar todos los parámetros de dispersión de modo preciso. Alternativamente, la Figura 2.2 sugiere una falta de información para c_{SY} cuando λ está incluido en el modelo. Por lo tanto, parece que no es necesario en estos datos incluir una estructura de covarianzas ancestral para efectos ambientales maternos permanentes combinados con la interacción toro \times año, dado que ambos efectos aleatorios se encuentran compitiendo por la información. Más aún, por sí misma la interacción toro \times año no parece tener impacto alguno en la reducción de la magnitud de r_{AoAm} .

2.3.4. Correlaciones de Spearman entre los valores de cría predichos

En el Cuadro 2.3 se muestran las correlaciones por rangos de Spearman entre los valores de cría predichos de los ocho modelos ajustados.

Cuadro 2.3. Coeficientes de correlación por rangos de Spearman calculados entre los valores de cría directos (encima de la diagonal) y, entre los valores de cría maternos (debajo de la diagonal) de los ocho modelos animales propuestos.

Modelos	I	E	G	GE	ISY	ESY	GSY	GESY
I	1	0.965	0.972	0.971	0.985	0.960	0.958	0.968
E	0.854	1	0.995	0.996	0.941	0.998	0.986	0.995
G	0.851	0.976	1	0.999	0.950	0.993	0.990	0.998
GE	0.845	0.978	0.999	1	0.948	0.994	0.990	0.998
ISY	0.965	0.922	0.918	0.913	1	0.943	0.952	0.950
ESY	0.842	0.999	0.975	0.977	0.920	1	0.990	0.996
GSY	0.818	0.968	0.992	0.991	0.910	0.972	1	0.994
GESY	0.835	0.976	0.998	0.999	0.911	0.977	0.994	1

En todos los casos, las correlaciones de Spearman entre los valores de cría directos fueron mayores a 0.94, evidenciando la ausencia de diferencias significativas entre los órdenes de mérito obtenidos con los distintos modelos. Al considerar las correlaciones de Spearman entre los valores de cría maternos, se encontraron diferencias en el ranking de los animales entre el modelo I (y el ISY) con el resto. Sin embargo, las correlaciones de Spearman entre E, G, GE, ESY, GSY y GESY no difirieron entre ellas. Estos resultados sugieren que en el caso de los valores de cría maternos, considerar la correlación entre los efectos ambientales maternos permanentes de dos generaciones adyacentes (λ), el efecto genético de abuela materna (a_n), o ambos efectos simultáneamente, tendría un efecto similar en las evaluaciones genéticas de los animales.

2.3.5. Recuperación de la variación aditiva total en los modelos con efectos maternos

La última columna del Cuadro 2.2 muestra los valores de h_T^2 para cada modelo. Las estimaciones de los modelos E, G, GE, ESY, GSY y GESY fueron mayores a aquellas de los modelos I e ISY, básicamente debido a la magnitud menos negativa del parámetro σ_{AoAm} . Los modelos E y ESY recuperaron la mayor proporción de la varianza aditiva total

2.3.6. Tópicos algorítmicos y de convergencia

Un tema a destacar es la eficacia del muestreo utilizado, el cual permitió obtener estimaciones de los parámetros de dispersión de los complejos modelos maternos con efectos ancestrales. Digno de mención es el muestreo individual con distintos grados de credibilidad de los componentes de (co)varianza aditivos a través de la distribución GIW, y el muestreo de λ mediante el uso del algoritmo de Chib (1993). La gran ventaja de este último es que no es necesario recurrir al algoritmo Metrópolis-Hastings por etapas dentro del muestreo de Gibbs, tal como propusieron originalmente Quintanilla et al. (1999) para el modelo E.

2.3.7. General

El objetivo de esta investigación fue detectar cuál entre modelos sobreparametrizados induce una menor magnitud estimada de r_{AoAm} pero con las mayores estimaciones posibles de h_o^2 y h_m^2 . Tal como se discutió en los apartados anteriores, la inclusión de la interacción toro \times año como un efecto aleatorio adicional en el modelo, produce una reducción en las estimaciones de h_o^2 y h_m^2 con una ligera disminución en la magnitud de r_{AoAm} . Por lo tanto, fue más ventajoso incluir, o bien los efectos genéticos de abuela materna, o bien considerar la correlación entre los efectos ambientales maternos permanentes de madres y sus hijas futuras madres. Probablemente, estos dos efectos son un reflejo de la relación negativa que se observa entre el registro fenotípico de peso al destete de una madre y los de su progenie para el mismo carácter (Kress y Burfening, 1972; Johnsson y Morant, 1984). En consecuencia, vacas que son expuestas a altos niveles de alimentación en etapas tempranas de su crecimiento ven comprometida su futura habilidad materna, a pesar de ser genéticamente superiores, por tener menor producción de leche producto del exceso de tejido graso en sus ubres (síndrome de la “ubre engrasada”). Este fenómeno puede ser analizado desde dos puntos de vista distintos. El primero de ellos, como un efecto genético aditivo de abuela materna (Willham, 1972), presentado aquí dentro de los

modelos G, GE, GSY, y GESY. El segundo enfoque es el de un efecto aleatorio ambiental recursivo que se va transmitiendo de generación en generación, por vía materna (Quintanilla et al., 1999), reflejado aquí en los modelos E, GE, ESY y GESY. De este modo, una abuela materna con capacidad superior de amamantar sobrealimenta a su hija, afectando la futura habilidad materna de esta última hembra a futuro. Sin embargo, la nieta de la primera vaca, a su vez, va a presentar una mayor habilidad materna, y, así sucesivamente se observa una alternancia entre generaciones. Dado que en el presente análisis la inclusión de λ (modelo E) resultó en una estimación menos negativa de r_{AoAm} con las mayores estimaciones de h_o^2 y h_m^2 , y que el modelo materno con efectos genéticos de abuela materna (G) incrementa el número de valores de cría a predecir, es recomendable el uso de E para mejorar los modelos de evaluación genética actual para el carácter peso al destete. Por otra parte, este ajuste se realiza con escaso esfuerzo computacional adicional, en comparación con el que debe realizarse para ajustar el MAM tradicional.

La covarianza madre-hijo (Koch 1972; Willham 1972; Foulley y Lefort 1978; y Bijma 2006) es igual a:

$$\text{Cov}(O, D) = \frac{1}{2} \sigma_{Ao}^2 + \frac{5}{4} \sigma_{AoAm} + \frac{1}{2} \sigma_{Am}^2 + \sigma_{EoEm} \quad [2.32]$$

El último término, σ_{EoEm} , representa “la covarianza ambiental entre los efectos directos de una madre, expresados en su propio fenotipo, y el efecto materno de la madre, expresado en el fenotipo de su cría” (Bijma, 2006, página 801). Contrariamente a lo que ocurre con la inclusión de λ en las ecuaciones del modelo mixto, donde se multiplica σ_{Em}^2 en todas las posiciones que relacionan al registro fenotípico de la madre con sus hijas que luego serán madres, la covarianza σ_{EoEm} debe incluirse en todas las posiciones que relacionan el registro de una madre con los datos de toda su progenie, tanto hembras como machos. Bijma (2006) observó que “las covarianzas ambientales entre las madres y sus crías tienden a ser un fenómeno general” en lugar de un “caso especial, como lo es el síndrome de la ubre engrasada en bovinos de carne”. Aún más, la naturaleza del proceso puede ser distinta en especies pluríparas si se las compara con las especies uníparas. En la investigación actual, σ_{EoEm} fue asumido igual a cero dado que en este archivo de datos hay un elevado número de vacas con más de una cría con registro fenotípico. En consecuencia, la covarianza entre el error (e_o) y los efectos ambientales maternos permanentes tendría una estructura muy compleja, no pudiéndose utilizar la metodología propuesta por Bijma (2006). Recientemente, Munilla Leguizamón y Cantet (2010) estimaron σ_{EoEm} en esta misma base de datos de la raza Angus.

Capítulo 3

Modelos animales con efectos maternos o de abuela materna equivalentes, cuando las madres y/o las abuelas maternas son desconocidas.

3.1. INTRODUCCIÓN

En ganado de carne existen ciertas razas o sub-poblaciones con políticas “abiertas” para registrar animales, a diferencia de las razas británicas tradicionales como Angus o Hereford, donde la inscripción al libro genealógico (herdbook) sólo puede producirse cuando ambos padres del animal se encuentran ya registrados. En las razas de pedigrí “abierto”, una fracción importante de animales con datos poseen madres que no tienen medición propia para los caracteres maternalmente afectados y, además, ninguno de sus padres (los abuelos maternos de la cría con dato) se encuentra identificado. Como consecuencia de dicho proceso, la evaluación genética de razas abiertas para caracteres con efectos maternos sufre de errores de especificación, dado que cuando un individuo con datos posee información faltante en relación con la madre y/o su abuela materna (Cantet et al., 1992b), la especificación de la estructura de covarianzas del modelo es distinta de la covarianza entre parientes con efectos maternos tal como la postulara Willham (1963) y fuera incorporada en el modelo animal por Quaas y Pollak (1980). En tales condiciones, los predictores de los valores de cría resultantes no son de mínima varianza (Henderson, 1975).

Cuando un individuo con datos para un carácter afectado maternalmente tiene madre y/o abuela materna desconocidas, existen distintas maneras de escribir el modelo de evaluación para conseguir la especificación apropiada de la estructura de covarianzas (Cantet et al., 1992b). Utilizar una u otra dependerá de las dificultades para calcular la matriz residual del modelo. Cuando la matriz de covarianzas de un modelo animal con efectos indirectos (maternos y/o de abuela materna) no es especificada correctamente, el modelo ajustado no explica totalmente la variación aditiva (Cantet et al., 1992b), y muchas covarianzas entre individuos no resultan iguales a las que se obtienen con el modelo de Willham (1972, fórmula [2]). Este problema es generalmente ignorado. Dado que para cualquier individuo con datos la matriz de covarianzas se puede especificar correctamente incluyendo la madre (Van Vleck, 1990a; Cantet et al., 1992b) y la abuela materna (Van Vleck, 1990a) en el vector de valores de cría, en caso que una, la otra o ambas sean desconocidas, pueden incluirse como individuos “fantasmas” (Van Vleck, 1990a; Cantet et al., 1992b). Sin embargo, mediante esta implementación aumenta el número de soluciones en el vector de valores de cría en una cantidad igual a dos veces el número de individuos fantasmas. Existen razas donde la fracción de animales con madre desconocida, o cuyas madres poseen esa única cría y no están emparentadas con cualquier otro individuo en el pedigrí (posiblemente por ser sus padres desconocidos), es elevada y a veces superior a un 30% del archivo de datos. En dichas condiciones, incluir madres y/o abuelas fantasmas en el vector \mathbf{a} es una solución computacionalmente inconveniente. Con el objeto de especificar correctamente la matriz de covarianzas de modelos animales con efectos maternos y/o de abuela materna cuando la madre y/o la abuela materna no se encuentran identificadas, en el presente capítulo se proponen modelos equivalentes alternativos a los presentados por Van Vleck (1990a), Cantet et al. (1992b) y Doderhoff et al. (1998, 1999 a,b). Cuando el vector de observaciones tiene el mismo valor esperado y la misma matriz de (co)varianzas ambos modelos son

equivalentes (Henderson, 1985), si bien en general uno de ellos tiene mayor facilidad de cómputo para ser ajustado que el otro.

3.2. ESPECIFICACIÓN DE LA COVARIANZA ADITIVA PARA REGISTROS DONDE LAS MADRES SON DESCONOCIDAS

Para especificar correctamente la porción aditiva de las varianzas y las covarianzas cuando faltan datos de las madres, y no tener que agregar una madre desconocida al pedigrí en el modelo con efectos maternos, R. L. Quaas a principio de los años 90, y en un intercambio epistolar con R. J. C. Cantet, sugirió expresar el valor de cría materno de una hembra desconocida cuyo hijo tiene un registro, como una función lineal del valor de cría de su progenie (multiplicado por $2/3$) y del padre de la cría (multiplicado por $-1/3$). Si el padre del animal es también desconocido, entonces el valor de cría materno de la hembra es reemplazado por la mitad del valor de cría materno de la progenie. Nótese que estas soluciones evitan tener que agregar en los vectores de valores de cría aquellos relacionados con las madres “fantasmas”. A continuación se reproducirá la demostración sobre la magnitud de los valores de reemplazo ($2/3$ y $-1/3$) realizada por R. J. C. Cantet a partir de la idea de R. L. Quaas.

Se desea obtener la predicción BLUP (*Best Lineal Unbiased Prediction*, predictor insesgado de mínima varianza, Henderson, 1984) del valor de cría de la j -ésima madre no identificada, en función de los BLUPs del individuo i y de su padre k de modo tal que

$$\text{BLUP}(a_j) = b_k \text{BLUP}(a_k) + b_i \text{BLUP}(a_i) \quad [3.1]$$

Los escalares b_k y b_i son los “coeficientes de regresión” de los $\text{BLUP}(a_k)$ y $\text{BLUP}(a_i)$, respectivamente, en el $\text{BLUP}(a_j)$. Ahora se demostrará que, cuando el padre k no está emparentado con la madre j y en ausencia de consanguinidad de k , $b_k = -1/3$ y $b_i = 2/3$.

El BLUP es el estimador de la esperanza condicional de la variable aleatoria u con respecto a la variable aleatoria x , e igual a (Henderson, 1984)

$$\text{BLUP}(u) = \hat{E}(u|x) = \text{cov}(u, x) [\text{var}(x)]^{-1} \text{BLUP}(x) \quad [3.2]$$

En nuestro caso, $u = a_j$, y $x = [a_k, a_i]'$, por lo tanto:

$$\text{BLUP}(a_j) = \text{cov} \left[a_j, \begin{bmatrix} a_k \\ a_i \end{bmatrix} \right] \left[\text{Var} \begin{bmatrix} a_k \\ a_i \end{bmatrix} \right]^{-1} \begin{bmatrix} \hat{a}_k \\ \hat{a}_i \end{bmatrix} = \begin{bmatrix} b_k & b_i \end{bmatrix} \begin{bmatrix} \hat{a}_k \\ \hat{a}_i \end{bmatrix}$$

donde el símbolo $\hat{}$ indica BLUP de la variable aleatoria. Ahora bien, por la teoría de la covarianza entre parientes en un modelo aditivo podemos indicar que:

$$\text{cov} \left[a_j, \begin{bmatrix} a_k \\ a_i \end{bmatrix} \right] = \begin{bmatrix} 0 \\ \frac{1}{2} \sigma_{Ao}^2 \end{bmatrix} \quad [3.3]$$

Como en capítulos anteriores, σ_{Ao}^2 es la varianza aditiva por efectos directos. Como j es una madre desconocida, debemos asumir que no está emparentada con el padre k , por lo cual $\text{cov}(a_j, a_k) = 0$. Asimismo

$$\text{Var} \begin{bmatrix} a_k \\ a_i \end{bmatrix} = \begin{bmatrix} 1 + F_k & \frac{1}{2}(1 + F_k) \\ \frac{1}{2}(1 + F_k) & 1 \end{bmatrix} \sigma_{Ao}^2$$

La inversa de esta matriz es:

$$\left[\text{Var} \begin{bmatrix} a_k \\ a_i \end{bmatrix} \right]^{-1} = \frac{1}{\sigma_{Ao}^2} \begin{bmatrix} \left[(1 + F_k) - \frac{1}{4}(1 + F_k)^2 \right]^{-1} & - \left[2 - \frac{(1 + F_k)}{2} \right]^{-1} \\ - \left[2 - \frac{(1 + F_k)}{2} \right]^{-1} & \frac{1}{\left[1 - \frac{1}{4}(1 + F_k) \right]} \end{bmatrix} \quad [3.4]$$

Premultiplicando [3.4] con [3.3], tenemos que:

$$\begin{bmatrix} b_k \\ b_i \end{bmatrix} = \text{cov} \left[a_j, \begin{bmatrix} a_k \\ a_i \end{bmatrix} \right] \text{Var} \begin{bmatrix} a_k \\ a_i \end{bmatrix}^{-1} = \begin{bmatrix} \frac{-1}{4 \left[1 - \frac{1}{4}(1 + F_k) \right]} \\ \frac{1}{2 \left[1 - \frac{1}{4}(1 + F_k) \right]} \end{bmatrix} = \begin{bmatrix} \frac{-1}{\left[3 - F_k \right]} \\ \frac{2}{\left[3 - F_k \right]} \end{bmatrix} \quad [3.5]$$

Cuando el padre del individuo no es consanguíneo $F_k = 0$; por lo tanto la expresión [3.5] resulta igual a:

$$\begin{bmatrix} b_k \\ b_i \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix}$$

Al ser aplicada en [3.1], esta solución produce entonces

$$\hat{a}_j = \frac{2}{3} \hat{a}_i - \frac{1}{3} \hat{a}_k \quad [3.6]$$

A pesar de que la expresión [3.6] permite especificar correctamente todas las (co)varianzas aditivas en los datos de individuos con sus madres no identificadas, existe una fracción de la varianza aditiva que no es explicada por el modelo cuando la ecuación del modelo animal con efectos maternos es igual a:

$$y_{ijk} = \mathbf{x}_i \boldsymbol{\beta} + a_{oi} + \frac{2}{3} a_{mi} - \frac{1}{3} a_{mk} + \epsilon_{oi} \quad [3.7]$$

Esto se debe a que:

$$\begin{aligned} \text{Var}(y_{ijk}) &= \text{Var}\left(a_{oi} + \frac{2}{3} a_{mi} - \frac{1}{3} a_{mk} + \epsilon_{oi}\right) \\ &= \text{Var}(a_{oi}) + \text{Var}\left(\frac{2}{3} a_{mi}\right) + \text{Var}\left(-\frac{1}{3} a_{mk}\right) + \text{Var}(\epsilon_{oi}) + 2\text{cov}\left(a_{oi}, \frac{2}{3} a_{mi}\right) + 2\text{cov}\left(a_{oi}, -\frac{1}{3} a_{mk}\right) + 2\text{cov}\left(\frac{2}{3} a_{mi}, -\frac{1}{3} a_{mk}\right) \end{aligned}$$

$$\begin{aligned}
&= \sigma_{Ao}^2 + \frac{4}{9}\sigma_{Am}^2 + \frac{1}{9}\sigma_{Am}^2 + \sigma_{Eo}^2 + 2\left(\frac{2}{3}\right)\sigma_{AoAm} + 2\left(-\frac{1}{3}\right)\frac{1}{2}\sigma_{AoAm} + 2\left(\frac{2}{3}\right)\left(-\frac{1}{3}\right)\frac{1}{2}\sigma_{Am}^2 \\
&= \sigma_{Ao}^2 + \frac{4}{9}\sigma_{Am}^2 + \frac{1}{9}\sigma_{Am}^2 + \sigma_{Eo}^2 + \frac{4}{3}\sigma_{AoAm} - \frac{1}{3}\sigma_{AoAm} - \frac{2}{9}\sigma_{Am}^2 \\
&= \sigma_{Ao}^2 + \left[\frac{4}{9} + \frac{1}{9} - \frac{2}{9}\right]\sigma_{Am}^2 + \left[\frac{4}{3} - \frac{1}{3}\right]\sigma_{AoAm} + \sigma_{Eo}^2 \\
&= \sigma_{Ao}^2 + \frac{1}{3}\sigma_{Am}^2 + \sigma_{AoAm} + \sigma_{Eo}^2
\end{aligned} \tag{3.8}$$

En consecuencia, es necesario agregar a la varianza del error la fracción de la varianza aditiva no contemplada por el modelo [3.7]. En este caso, al error puro se le debe adicionar $\frac{2}{3}\sigma_{Am}^2 + \sigma_{Em}^2$, resultando la varianza de ϵ_{oi} igual a:

$$\text{Var}(\epsilon_{oi}) = \sigma_{Eo}^2 + \frac{2}{3}\sigma_{Am}^2 + \sigma_{Em}^2 \tag{3.9}$$

Cuando ambos padres son desconocidos, en cambio, el valor de cría materno de la hembra j se reemplaza por el valor de cría de su progenie i ,

$$a_{mj} = \frac{1}{2}a_{mi} \tag{3.10}$$

y la ecuación del animal (Quaas, 1992) es igual a:

$$y_{ijk} = \mathbf{x}_i \boldsymbol{\beta} + a_{oi} + \frac{1}{2}a_{mi} + \epsilon_{oi} \tag{3.11}$$

Al tomar el operador varianza en [3.11], tenemos:

$$\begin{aligned}
\text{Var}(y_{ijk}) &= \text{Var}(a_{oi}) + \text{Var}\left(\frac{1}{2}a_{mi}\right) + \text{Var}(\epsilon_{oi}) + 2\text{cov}\left(a_{oi}, \frac{1}{2}a_{mi}\right) \\
&= \sigma_{Ao}^2 + \frac{1}{4}\sigma_{Am}^2 + \sigma_{Eo}^2 + 2\left(\frac{1}{2}\right)\sigma_{AoAm} \\
&= \sigma_{Ao}^2 + \frac{1}{4}\sigma_{Am}^2 + \sigma_{Eo}^2 + \sigma_{AoAm}
\end{aligned} \tag{3.12}$$

Por lo tanto, en este caso, al error puro se le debe adicionar $\frac{3}{4}\sigma_{Am}^2 + \sigma_{Em}^2$, resultando entonces la varianza de ϵ_{oi} igual a:

$$\text{Var}(\epsilon_{oi}) = \sigma_{Eo}^2 + \frac{3}{4}\sigma_{Am}^2 + \sigma_{Em}^2 \tag{3.13}$$

La ponderación de a_{mj} surge de:

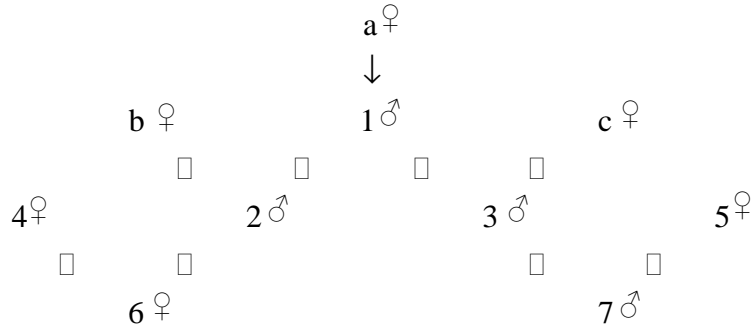
$$b = \frac{\text{cov}(a_i, a_j)}{\text{Var}(a_i)} = \frac{1}{2} \frac{\sigma_{Ao}^2}{\sigma_{Ao}^2} = \frac{1}{2}$$

Dicha expresión es la que explica la inclusión del valor de cría materno del individuo dentro del modelo para su propio registro fenotípico.

A continuación se presenta un pequeño ejemplo numérico con el objetivo de evidenciar la equivalencia entre el modelo propuesto por R. L. Quaas y el MAM convencional.

3.2.1. Ejemplo numérico

Tomando los números por individuos conocidos y las letras por desconocidos, suponga el siguiente pedigrí:



Asumiremos ahora que los individuos 1, 2, 3, 6 y 7 poseen registros fenotípicos para el carácter peso al destete, siendo estos datos iguales a $\mathbf{y}' = [150 \ 133 \ 128 \ 146 \ 120]$, respectivamente. Las madres a, b y c son individuos “fantasmas” (Van Vleck, 1990a). Se considera la edad al destete, expresada en días, como el único efecto fijo (β_1). El vector de incidencia para β_1 es $\mathbf{x}' = [180 \ 161 \ 151 \ 162 \ 132]$. Los componentes de (co)varianza se asumen iguales a $\mathbf{G}_o = \begin{bmatrix} 100 & -25 \\ -25 & 75 \end{bmatrix}$, $\sigma_{Em}^2 = 50$ y $\sigma_{Eo}^2 = 500$. Al ajustar el MAM tradicional, se consideran los siguientes vectores de valores de cría directos

$$\mathbf{a}_o' = [a_{oa} \ a_{ob} \ a_{oc} \ a_{o1} \ a_{o2} \ a_{o3} \ a_{o4} \ a_{o5} \ a_{o6} \ a_{o7}] ,$$

valores de cría maternos

$$\mathbf{a}_m' = [a_{ma} \ a_{mb} \ a_{mc} \ a_{m1} \ a_{m2} \ a_{m3} \ a_{m4} \ a_{m5} \ a_{m6} \ a_{m7}] ,$$

y efectos ambientales maternos permanentes

$$\mathbf{e}_m' = [e_{ma} \ e_{mb} \ e_{mc} \ e_{m4} \ e_{m5}]$$

con sus respectivas matrices de incidencia:

$$\mathbf{Z}_o = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \mathbf{Z}_m = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}; \mathbf{Z}_p = \mathbf{I}_5 \quad [3.14]$$

Se asume que la matriz de (co)varianzas para el vector \mathbf{a} es $\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_o$, donde la matriz \mathbf{A} de relaciones de parentesco correspondiente al pedigrí analizado es:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0.50 & 0.25 & 0.25 & 0 & 0 & 0.125 & 0.125 \\ 0 & 1 & 0 & 0 & 0.50 & 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.50 & 0 & 0 & 0 & 0.25 \\ 0.50 & 0 & 0 & 1 & 0.50 & 0.50 & 0 & 0 & 0.25 & 0.25 \\ 0.25 & 0.50 & 0 & 0.50 & 1 & 0.25 & 0 & 0 & 0.50 & 0.125 \\ 0.25 & 0 & 0.50 & 0.50 & 0.25 & 1 & 0 & 0 & 0.125 & 0.50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0.50 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0.50 \\ 0.125 & 0.25 & 0 & 0.25 & 0.50 & 0.125 & 0.50 & 0 & 1 & 0.0625 \\ 0.125 & 0 & 0.25 & 0.25 & 0.125 & 0.50 & 0 & 0.50 & 0.0625 & 1 \end{bmatrix} \quad [3.15]$$

La matriz de (co)varianzas para \mathbf{e}_m es $\mathbf{I}_5 \sigma_{Em}^2$. Finalmente, se asume que las 5 observaciones tienen el mismo error residual, y que dichos errores no están correlacionados, por lo tanto, la matriz de (co)varianzas del vector de errores es $\mathbf{R} = \mathbf{I}_5 \sigma_{Eo}^2$.

Las ecuaciones de modelo mixto son iguales a:

$$\begin{bmatrix} \mathbf{x}'\mathbf{R}^{-1}\mathbf{x} & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_o & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_m & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_o + \mathbf{A}^{-1}\frac{3}{275} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_m + \mathbf{A}^{-1}\frac{1}{275} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_o + \mathbf{A}^{-1}\frac{1}{275} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_m + \mathbf{A}^{-1}\frac{4}{275} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_o & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_m & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_p + \mathbf{I}_5\frac{1}{50} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\mathbf{a}}_o \\ \hat{\mathbf{a}}_m \\ \hat{\mathbf{e}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [3.16]$$

Al reemplazar [3.14], [3.15] y la inversa de \mathbf{R} en [3.16] y resolver el sistema de ecuaciones, se obtienen las siguientes soluciones:

$$\hat{\beta}_1 = 0.859; \quad \begin{bmatrix} \hat{a}_{oa} \\ \hat{a}_{ob} \\ \hat{a}_{oc} \\ \hat{a}_{o1} \\ \hat{a}_{o2} \\ \hat{a}_{o3} \\ \hat{a}_{o4} \\ \hat{a}_{o5} \\ \hat{a}_{o6} \\ \hat{a}_{o7} \end{bmatrix} = \begin{bmatrix} -0.191 \\ 0.053 \\ 0.174 \\ -0.637 \\ -0.440 \\ -0.125 \\ 0.255 \\ 0.242 \\ 0.419 \\ 0.543 \end{bmatrix}; \quad \begin{bmatrix} \hat{a}_{ma} \\ \hat{a}_{mb} \\ \hat{a}_{mc} \\ \hat{a}_{m1} \\ \hat{a}_{m2} \\ \hat{a}_{m3} \\ \hat{a}_{m4} \\ \hat{a}_{m5} \\ \hat{a}_{m6} \\ \hat{a}_{m7} \end{bmatrix} = \begin{bmatrix} -0.419 \\ -0.569 \\ -0.231 \\ -0.074 \\ -0.284 \\ -0.179 \\ 0.639 \\ 0.606 \\ 0.049 \\ 0.092 \end{bmatrix}; \quad \begin{bmatrix} \hat{e}_{ma} \\ \hat{e}_{mb} \\ \hat{e}_{mc} \\ \hat{e}_{m4} \\ \hat{e}_{m5} \end{bmatrix} = \begin{bmatrix} -0.339 \\ -0.404 \\ -0.136 \\ 0.511 \\ 0.485 \end{bmatrix} \quad [3.17]$$

Ahora suponga que se desea reemplazar el valor de cría materno de a por la mitad del valor de cría de 1, y expresar los valores de cría maternos de los individuos b, c, 4 y 5 en función de los de sus crías y de los padres de sus crías como en el modelo [3.7]. De este modo, se reduce notablemente el número de individuos a tener en cuenta en los vectores \mathbf{a}_o , \mathbf{a}_m y \mathbf{e}_m , con la consecuente reducción en el número de ecuaciones a resolver. En este ejemplo, al no incluir las evaluaciones de los animales a, b, c, 4 y 5, se pasa de 26 ecuaciones a 11, es decir, se resolvieron solamente el 42.3 % de las incógnitas. Aquí, los vectores de valores de cría son: $\mathbf{a}_o' = [a_{o1} \ a_{o2} \ a_{o3} \ a_{o6} \ a_{o7}]$ y $\mathbf{a}_m' = [a_{m1} \ a_{m2} \ a_{m3} \ a_{m6} \ a_{m7}]$. Por su parte, las matrices de incidencia para los efectos directos y maternos son iguales a:

$$\mathbf{Z}_o = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{Z}_m = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & -\frac{1}{3} & 0 & \frac{2}{3} \end{bmatrix} \quad [3.18]$$

Adicionalmente, de la matriz de relaciones presentada en [3.15] se deben eliminar las filas y columnas correspondientes a los animales a, b, c, 4 y 5, resultando la matriz \mathbf{A} reducida igual a:

$$\mathbf{A} = \begin{bmatrix} 1 & 0.50 & 0.50 & 0.25 & 0.25 \\ 0.50 & 1 & 0.25 & 0.50 & 0.125 \\ 0.50 & 0.25 & 1 & 0.125 & 0.50 \\ 0.25 & 0.50 & 0.125 & 1 & 0.0625 \\ 0.25 & 0.125 & 0.50 & 0.0625 & 1 \end{bmatrix} \quad [3.19]$$

A diferencia de lo que ocurre en el MAM, bajo este modelo, al elemento diagonal de la matriz de (co)varianzas del efecto residual del individuo 1 se le debe sumar

$\frac{3}{4}(75) + 50$ (expresión [3.13]), mientras que a los individuos 2, 3, 6 y 7 se les debe adicionar $\frac{2}{3}(75) + 50$ (expresión [3.9]) en sus elementos diagonales, además de la varianza del error. En consecuencia, la matriz \mathbf{R} resultante es diagonal e igual a:

$$\mathbf{R} = \begin{bmatrix} 606.25 & 0 & 0 & 0 & 0 \\ 0 & 600 & 0 & 0 & 0 \\ 0 & 0 & 600 & 0 & 0 \\ 0 & 0 & 0 & 600 & 0 \\ 0 & 0 & 0 & 0 & 600 \end{bmatrix} \quad [3.20]$$

Las ecuaciones de modelo mixto a resolver son entonces iguales a:

$$\begin{bmatrix} \mathbf{x}'\mathbf{R}^{-1}\mathbf{x} & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_o & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_m \\ \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_o + \mathbf{A}^{-1}\frac{3}{275} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_m + \mathbf{A}^{-1}\frac{1}{275} \\ \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_o + \mathbf{A}^{-1}\frac{1}{275} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_m + \mathbf{A}^{-1}\frac{4}{275} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\mathbf{a}}_o \\ \hat{\mathbf{a}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [3.21]$$

Nótese que, por las dimensiones del ejemplo, no se pudieron ajustar efectos ambientales maternos permanentes. Sin embargo, si se hubieran tenido en cuenta animales con datos cuyas madres fueran conocidas, se hubieran resuelto las ecuaciones de modelo mixto presentadas en el MAM convencional (ecuación [3.16]). Por último, al resolver [3.21], con [3.18], [3.19] y la inversa de [3.20], se obtienen las siguientes soluciones:

$$\hat{\beta}_1 = 0.859; \quad \begin{bmatrix} \hat{a}_{o1} \\ \hat{a}_{o2} \\ \hat{a}_{o3} \\ \hat{a}_{o6} \\ \hat{a}_{o7} \end{bmatrix} = \begin{bmatrix} -0.637 \\ -0.440 \\ -0.125 \\ 0.419 \\ 0.543 \end{bmatrix}; \quad \begin{bmatrix} \hat{a}_{m1} \\ \hat{a}_{m2} \\ \hat{a}_{m3} \\ \hat{a}_{m6} \\ \hat{a}_{m7} \end{bmatrix} = \begin{bmatrix} -0.074 \\ -0.284 \\ -0.179 \\ 0.049 \\ 0.092 \end{bmatrix}; \quad [3.22]$$

Nótese que las soluciones obtenidas en [3.22] son iguales a las obtenidas en [3.17] para los individuos 1, 2, 3, 6 y 7. Esto enfatiza la equivalencia de modelos. Así, la estimación de la esperanza del vector de observaciones es la misma en ambos modelos, e igual a:

$$\mathbf{E}(\mathbf{y}) = \mathbf{x} \hat{\beta}_1 = \begin{bmatrix} 180 \\ 161 \\ 151 \\ 162 \\ 132 \end{bmatrix} 0.859 = \begin{bmatrix} 154.62 \\ 138.30 \\ 129.71 \\ 139.16 \\ 113.39 \end{bmatrix} \quad [3.23]$$

Asimismo, la matriz de (co)varianzas de las cinco observaciones (V) es:

$$V = \begin{bmatrix} 700 & 43.75 & 43.75 & 21.875 & 21.875 \\ 43.75 & 700 & 25 & 43.75 & 12.50 \\ 43.75 & 25 & 700 & 12.50 & 43.75 \\ 21.875 & 43.75 & 12.50 & 700 & 6.25 \\ 21.875 & 12.50 & 43.75 & 6.25 & 700 \end{bmatrix} \quad [3.24]$$

La expresión [3.24] se obtuvo mediante:

a) $V = Z_o A Z_o' 100 - 25(Z_o A Z_m' + Z_m A Z_o') + Z_m A Z_m' 75 + Z_p Z_p' 50 + I_5 500$ en el MAM convencional; y

b) $V = Z_o A Z_o' 100 - 25(Z_o A Z_m' + Z_m A Z_o') + Z_m A Z_m' 75 + R$, en el modelo de R. L. Quaas.

3.3. ESPECIFICACIÓN DE LA COVARIANZA ADITIVA PARA REGISTROS DONDE LAS ABUELAS MATERNAS SON DESCONOCIDAS

Trasladaremos el razonamiento empleado por R. L. Quaas, presentado en la sección anterior, al caso en que la abuela materna es desconocida y la madre conocida. En esencia, el procedimiento consiste en que, condicional al valor de cría de sus padres el residuo mendeliano de cualquier individuo es independiente de los valores de cría de todo otro animal (Bulmer, 1985) y por lo tanto la varianza residual del dato es diagonal. Esto posibilita calcular eficientemente las ecuaciones de modelo mixto (Schaeffer y Henderson, 1983). En este modelo, el valor de cría de abuela materna de los individuos cuya abuela materna no se encuentra identificada se debe reemplazar por:

$$a_{ng} = \frac{2}{3} a_{nj} - \frac{1}{3} a_{nl} \quad [3.25]$$

o

$$a_{ng} = \frac{1}{2} a_{nj} \quad [3.26]$$

donde l es el abuelo materno del individuo i . Sin embargo, a pesar de que las expresiones [3.25] y [3.26] permiten especificar correctamente todas las covarianzas aditivas en los datos de individuos con abuelas maternas no identificadas, existe una fracción de la varianza aditiva que no es explicada por el modelo y que es necesario entonces agregar a la varianza del error.

Cuando la abuela materna de i es desconocida y su valor de cría predicho es reemplazado por el BLUP de los valores de cría de su hija y del padre de su progenie (l), tal como lo indica la expresión [3.25], al error puro se le debe adicionar $\frac{2}{3} \sigma_{An}^2$, es decir las dos terceras partes de la varianza aditiva de los efectos de abuela materna. Por lo tanto, la varianza del error (ϵ_{oi}) resulta igual a:

$$\text{Var}(\epsilon_{oi}) = \sigma_{Eo}^2 + \frac{2}{3} \sigma_{An}^2 \quad [3.27]$$

En cambio, cuando el valor de cría de la abuela materna g es reemplazado por $\frac{1}{2}a_{nj}$, la variabilidad aditiva que va a parar al error es $\frac{3}{4}\sigma_{An}^2$, siendo $\text{Var}(\epsilon_{oi})$ igual a

$$\text{Var}(\epsilon_{oi}) = \sigma_{Eo}^2 + \frac{3}{4}\sigma_{An}^2 \quad [3.28]$$

Las expresiones [3.27] y [3.28] son válidas siempre y cuando se asuma que el abuelo l no es consanguíneo. Si esto no fuese así, y tomando F_l como el coeficiente de consanguinidad de l , la fracción de varianza aditiva en el error es inferior y será proporcional a:

$$\frac{1}{3}(2 - F_l)\sigma_{An}^2 \quad [3.29]$$

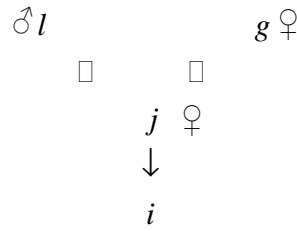
o

$$\frac{1}{4}(3 - F_l)\sigma_{An}^2 \quad [3.30]$$

En la siguiente sección se detallan las derivaciones necesarias para llegar a los resultados obtenidos en [3.27] y [3.28].

3.3.1. Obtención de la varianza de los ϵ_{oi}

En el siguiente pedigrí, l representa un macho abuelo materno de i , g la abuela materna de i , j es la madre de i , e i es el individuo con datos.



A partir de la ecuación escalar del modelo animal con efectos de abuela materna

$$y_{ijg} = \mathbf{x}_i \boldsymbol{\beta} + a_{oi} + a_{mj} + a_{ng} + e_{mj} + e_{oi} \quad [3.31]$$

la varianza de una observación es igual a

$$\begin{aligned}
 \text{Var}(y_{ijk}) &= \text{Var}(a_{oi} + a_{mj} + a_{ng} + e_{mj} + e_{oi}) \\
 &= \text{Var}(a_{oi}) + \text{Var}(a_{mj}) + \text{Var}(a_{ng}) + \text{Var}(e_{mj}) + \text{Var}(e_{oi}) + 2\text{cov}(a_{oi}, a_{mj}) \\
 &\quad + 2\text{cov}(a_{oi}, a_{ng}) + 2\text{cov}(a_{mj}, a_{ng}) \\
 &= (1 + F_i)\sigma_{Ao}^2 + (1 + F_j)\sigma_{Am}^2 + (1 + F_g)\sigma_{An}^2 + \sigma_{Em}^2 + \sigma_{Eo}^2 + 2\left[F_i + \frac{1}{2}(1 + F_j)\right]\sigma_{AoAm} \\
 &\quad + 2\left[F_i + \frac{1}{4}(1 + F_g)\right]\sigma_{AoAn} + 2\left[F_j + \frac{1}{2}(1 + F_g)\right]\sigma_{AmAn}
 \end{aligned}$$

En los casos que se desconoce la abuela, resulta adecuado asumir que $F_g = F_j = 0$.

$$\begin{aligned}
 &= (1 + F_i) \sigma_{Ao}^2 + (1 + F_j) \sigma_{Am}^2 + (1 + F_g) \sigma_{An}^2 + \sigma_{Em}^2 + \sigma_{Eo}^2 + 2 \left(\frac{1}{2} \right) \sigma_{AoAm} + 2 \left(\frac{1}{4} \right) \sigma_{AoAn} + 2 \left(\frac{1}{2} \right) \sigma_{AmAn} \\
 &= (1 + F_i) \sigma_{Ao}^2 + (1 + F_j) \sigma_{Am}^2 + (1 + F_g) \sigma_{An}^2 + \sigma_{AoAm} + \frac{1}{2} \sigma_{AoAn} + \sigma_{AmAn} + \sigma_{Em}^2 + \sigma_{Eo}^2 \quad [3.32]
 \end{aligned}$$

Cuando la abuela materna g es desconocida, su valor de cría para los efectos de abuela materna es reemplazado por una función lineal de los valores de cría de abuela materna de su hija j y del padre de su hija l de modo tal que:

$$y_{ijl} = \mathbf{x}_i \boldsymbol{\beta} + a_{oi} + a_{mj} + \frac{2}{3} a_{nj} - \frac{1}{3} a_{nl} + e_{mj} + e_{oi} \quad [3.33]$$

Si se toma el operador varianza en [3.33] se obtiene:

$$\begin{aligned}
 \text{Var}(y_{ijl}) &= \text{Var}\left(a_{oi} + a_{mj} + \frac{2}{3} a_{nj} - \frac{1}{3} a_{nl} + e_{mj} + \epsilon_{oi}\right) \\
 &= \text{Var}(a_{oi}) + \text{Var}(a_{mj}) + \frac{4}{9} \text{Var}(a_{nj}) + \frac{1}{9} \text{Var}(a_{nl}) + \text{Var}(e_{mj}) + \text{Var}(\epsilon_{oi}) \\
 &\quad + 2 \text{cov}(a_{oi}, a_{mj}) + 2 \text{cov}\left(a_{oi}, \frac{2}{3} a_{nj}\right) + 2 \text{cov}\left(a_{oi}, -\frac{1}{3} a_{nl}\right) \\
 &\quad + 2 \text{cov}\left(a_{mj}, \frac{2}{3} a_{nj}\right) + 2 \text{cov}\left(a_{mj}, -\frac{1}{3} a_{nl}\right) + 2 \text{cov}\left(\frac{2}{3} a_{nj}, -\frac{1}{3} a_{nl}\right) \\
 &= (1 + F_i) \sigma_{Ao}^2 + (1 + F_j) \sigma_{Am}^2 + \frac{4}{9} (1 + F_j) \sigma_{An}^2 + \frac{1}{9} (1 + F_l) \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) + 2 \left(\frac{1}{2} \right) \sigma_{AoAm} \\
 &\quad + 2 \left(\frac{2}{3} \right) \frac{1}{2} \sigma_{AoAn} + 2 \left(-\frac{1}{3} \right) \frac{1}{4} \sigma_{AoAn} + 2 \left(\frac{2}{3} \right) (1 + F_j) \sigma_{AmAn} + 2 \left(-\frac{1}{3} \right) \frac{1}{2} \sigma_{AmAn} + 2 \left(-\frac{1}{3} \right) \frac{1}{2} \sigma_{An}^2 \\
 &= (1 + F_i) \sigma_{Ao}^2 + (1 + F_j) \sigma_{Am}^2 + \frac{4}{9} (1 + F_j) \sigma_{An}^2 + \frac{1}{9} (1 + F_l) \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) \\
 &\quad + \sigma_{AoAm} + \frac{2}{3} \sigma_{AoAn} - \frac{1}{6} \sigma_{AoAn} + \frac{4}{3} (1 + F_j) \sigma_{AmAn} - \frac{1}{3} \sigma_{AmAn} - \frac{2}{9} \sigma_{An}^2 \\
 &= (1 + F_i) \sigma_{Ao}^2 + (1 + F_j) \sigma_{Am}^2 + \left[\frac{4}{9} (1 + F_j) + \frac{1}{9} (1 + F_l) - \frac{2}{9} \right] \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) \\
 &\quad + \sigma_{AoAm} + \left[\frac{2}{3} - \frac{1}{6} \right] \sigma_{AoAn} + \left[\frac{4}{3} (1 + F_j) - \frac{1}{3} \right] \sigma_{AmAn} \\
 &= (1 + F_i) \sigma_{Ao}^2 + (1 + F_j) \sigma_{Am}^2 + \left[\frac{1}{3} + \frac{(4F_j + F_l)}{9} \right] \sigma_{An}^2 + \sigma_{Em}^2 + \\
 &\quad \text{Var}(\epsilon_{oi}) + \sigma_{AoAm} + \frac{1}{2} \sigma_{AoAn} + \left(1 + \frac{4}{3} F_j \right) \sigma_{AmAn}
 \end{aligned}$$

Dado que g es desconocida, asumiremos que la madre de i no es consanguínea: $F_j = 0$. Consecuentemente, reemplazando en el último renglón de la expresión anterior obtenemos la varianza de una observación bajo el modelo [3.33], que resulta igual a:

$$\text{Var}(y_{ijl}) = (1 + F_i) \sigma_{Ao}^2 + \sigma_{Am}^2 + \left[\frac{1}{3} + \frac{F_l}{9} \right] \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) + \sigma_{AoAm} + \frac{1}{2} \sigma_{AoAn} + \sigma_{AmAn} \quad [3.34]$$

Para obtener la fracción de la varianza aditiva no explicada por el modelo [3.33], debemos obtener la diferencia entre [3.32] y [3.34] y despejar la $\text{Var}(\epsilon_{oi})$, de modo tal que el modelo resulte equivalente al de Willham (1972). A los efectos de simplificar la presentación suponga que el abuelo materno l no es consanguíneo (es decir, $F_l = 0$), entonces:

$$\text{Var}(\epsilon_{oi}) = \sigma_{Eo}^2 + \frac{2}{3} \sigma_{An}^2 \quad [3.35]$$

Existen otras situaciones en donde no sólo se desconoce a la abuela materna, sino que tampoco se conoce el abuelo materno. En estos casos, un modelo equivalente al de Willham (1972) es el siguiente:

$$y_{ij} = \mathbf{x}_i \boldsymbol{\beta} + a_{oi} + a_{mj} + \frac{1}{2} a_{nj} + e_{mj} + \epsilon_{oi} \quad [3.36]$$

Tomando el operador varianza en [3.36], se obtiene:

$$\begin{aligned} \text{Var}(y_{ij}) &= \text{Var}\left(a_{oi} + a_{mj} + \frac{1}{2} a_{nj} + e_{mj} + \epsilon_{oi}\right) \\ &= \text{Var}(a_{oi}) + \text{Var}(a_{mj}) + \frac{1}{4} \text{Var}(a_{nj}) + \text{Var}(e_{mj}) + \text{Var}(\epsilon_{oi}) \\ &\quad + 2 \text{cov}(a_{oi}, a_{mj}) + 2 \text{cov}\left(a_{oi}, \frac{1}{2} a_{nj}\right) + 2 \text{cov}\left(a_{mj}, \frac{1}{2} a_{nj}\right) \\ &= (1 + F_i) \sigma_{Ao}^2 + \sigma_{Am}^2 + \frac{1}{4} \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) + 2\left(\frac{1}{2}\right) \sigma_{AoAm} + 2\left(\frac{1}{2}\right) \text{cov}(a_{oi}, a_{mj}) + 2\left(\frac{1}{2}\right) \text{cov}(a_{mj}, a_{nj}) \\ &= (1 + F_i) \sigma_{Ao}^2 + \sigma_{Am}^2 + \frac{1}{4} \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) + \sigma_{AoAm} + \text{cov}(a_{oi}, a_{nj}) + \text{cov}(a_{mj}, a_{nj}) \\ &= (1 + F_i) \sigma_{Ao}^2 + \sigma_{Am}^2 + \frac{1}{4} \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) + \sigma_{AoAm} + \frac{1}{2} \sigma_{AoAn} + \text{cov}(a_{mj}, a_{nj}) \\ &= (1 + F_i) \sigma_{Ao}^2 + \sigma_{Am}^2 + \frac{1}{4} \sigma_{An}^2 + \sigma_{Em}^2 + \text{Var}(\epsilon_{oi}) + \sigma_{AoAm} + \frac{1}{2} \sigma_{AoAn} + \sigma_{AmAn} \end{aligned} \quad [3.37]$$

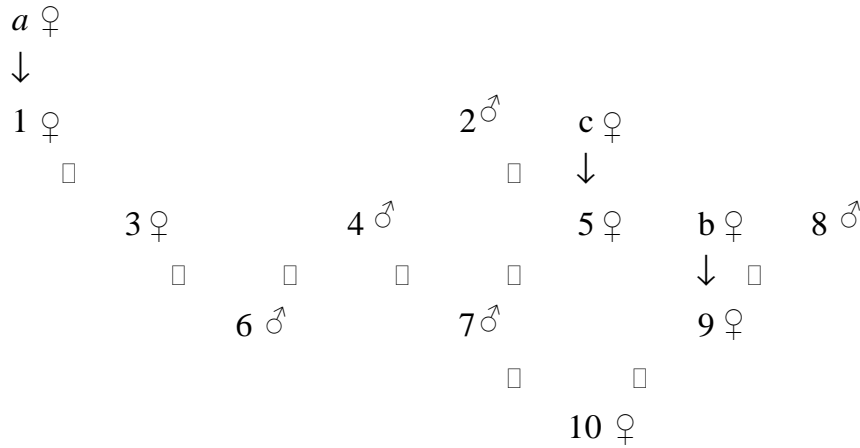
Nuevamente, como se procedió anteriormente, restando [3.37] a [3.32], obtenemos

$$\text{Var}(\epsilon_{oi}) = \sigma_{Eo}^2 + \frac{3}{4} \sigma_{An}^2 \quad [3.38]$$

Expresiones como [3.35] y [3.38] muestran la heterogeneidad de la varianza residual debida a la falta de información en el pedigrí. Nótese que la especificación de la covarianza entre parientes en el modelo con efectos maternos y de abuela materna (Willham, 1972) cuando la abuela del individuo con datos es desconocida se logra regresando apropiadamente los valores de cría de individuos conocidos en el pedigrí y adicionando al error una fracción de la varianza aditiva entre efectos de abuela materna.

3.3.2. Un pequeño ejemplo numérico

Desarrollaremos un ejemplo numérico para mostrar la equivalencia entre el modelo animal con efectos de abuela materna de Willham (1972) y el modelo aquí sugerido. Suponga el siguiente pedigrí:



Las hembras a, b y c son abuelas desconocidas agregadas (“fantasmas”). Los individuos 3, 6, 7 y 10 poseen dato de peso al destete. El vector de observaciones de los cuatro animales es $\mathbf{y}' = [150 \ 133 \ 128 \ 146]$. Se consideró la edad al destete, expresada en días, como el único efecto fijo (β_1). El vector de incidencia para β_1 es igual a $\mathbf{x}' = [180 \ 161 \ 151 \ 162]$. Los componentes de (co)varianza se asumen iguales a

$$\mathbf{G}_o = \begin{bmatrix} 100 & -25 & 15 \\ -25 & 75 & 18 \\ 15 & 18 & 60 \end{bmatrix}, \quad \sigma_{Em}^2 = 50 \text{ y } \sigma_{Eo}^2 = 500. \text{ Al ajustar el modelo animal con efectos de}$$

abuela materna de Willham (1972), se obtienen los siguientes vectores de valores de cría directos, maternos y de abuela materna, y de efectos ambientales maternos permanentes:

$$\begin{aligned} \mathbf{a}_o' &= [a_{oa} \ a_{ob} \ a_{oc} \ a_{o1} \ a_{o2} \ a_{o3} \ a_{o4} \ a_{o5} \ a_{o6} \ a_{o7} \ a_{o8} \ a_{o9} \ a_{o10}]; \\ \mathbf{a}_m' &= [a_{ma} \ a_{mb} \ a_{mc} \ a_{m1} \ a_{m2} \ a_{m3} \ a_{m4} \ a_{m5} \ a_{m6} \ a_{m7} \ a_{m8} \ a_{m9} \ a_{m10}]; \\ \mathbf{a}_n' &= [a_{na} \ a_{nb} \ a_{nc} \ a_{n1} \ a_{n2} \ a_{n3} \ a_{n4} \ a_{n5} \ a_{n6} \ a_{n7} \ a_{n8} \ a_{n9} \ a_{n10}]; \\ \mathbf{e}_m' &= [e_{m1} \ e_{m3} \ e_{m5} \ e_{m9}] \end{aligned}$$

Las matrices de incidencia de los efectos aleatorios son las siguientes:

$$\begin{aligned}
\mathbf{Z}_o &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{Z}_m = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \\
\mathbf{Z}_n &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad \mathbf{Z}_p = \mathbf{I}_4;
\end{aligned} \tag{3.39}$$

La matriz de (co)varianzas para el vector \mathbf{a} es $\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0$, donde \mathbf{A} es:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0.50 & 0 & 0.25 & 0 & 0 & 0.125 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & 0.25 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0.50 & 0 & 0.25 & 0 & 0 & 0.125 \\ 0.50 & 0 & 0 & 1 & 0 & 0.50 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0.50 & 0 & 0.25 & 0 & 0 & 0.125 \\ 0.25 & 0 & 0 & 0.50 & 0 & 1 & 0 & 0 & 0.50 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0.50 & 0.50 & 0 & 0 & 0.25 \\ 0 & 0 & 0.50 & 0 & 0.50 & 0 & 0 & 1 & 0 & 0.50 & 0 & 0 & 0.25 \\ 0.125 & 0 & 0 & 0.25 & 0 & 0.50 & 0.50 & 0 & 1 & 0.25 & 0 & 0 & 0.125 \\ 0 & 0 & 0.25 & 0 & 0.25 & 0 & 0.50 & 0.50 & 0.25 & 1 & 0 & 0 & 0.50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.50 & 0.25 \\ 0 & 0.50 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & 1 & 0.50 \\ 0 & 0.25 & 0.125 & 0 & 0.125 & 0 & 0.25 & 0.25 & 0.125 & 0.50 & 0.25 & 0.50 & 1 \end{bmatrix} \tag{3.40}$$

Los efectos ambientales maternos permanentes se asumen independientes entre sí:
 $\text{Var}(\mathbf{e}_m) = \mathbf{I}_4 \sigma_{\text{Em}}^2$. Por último, se asume que $\mathbf{R} = \mathbf{I}_4 \sigma_{\text{Eo}}^2$.

Las MME son:

$$\begin{bmatrix} \mathbf{x}'\mathbf{R}^{-1}\mathbf{x} & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_o & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_m & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_n & \mathbf{x}'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{11} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{12} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{13} & \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{21} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{22} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{23} & \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_n'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_n'\mathbf{R}^{-1}\mathbf{Z}_o + \mathbf{A}^{-1}\mathbf{g}^{31} & \mathbf{Z}_n'\mathbf{R}^{-1}\mathbf{Z}_m + \mathbf{A}^{-1}\mathbf{g}^{32} & \mathbf{Z}_n'\mathbf{R}^{-1}\mathbf{Z}_n + \mathbf{A}^{-1}\mathbf{g}^{33} & \mathbf{Z}_n'\mathbf{R}^{-1}\mathbf{Z}_p \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{x} & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_o & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_m & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_n & \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{Z}_p + \mathbf{I}_4 \frac{1}{50} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\mathbf{a}}_o \\ \hat{\mathbf{a}}_m \\ \hat{\mathbf{a}}_n \\ \hat{\mathbf{e}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_o'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_m'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_n'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_p'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{3.41}$$

$$\text{donde } \mathbf{G}_o^{-1} = \begin{bmatrix} \mathbf{g}^{11} & \mathbf{g}^{12} & \mathbf{g}^{13} \\ \mathbf{g}^{21} & \mathbf{g}^{22} & \mathbf{g}^{23} \\ \mathbf{g}^{31} & \mathbf{g}^{32} & \mathbf{g}^{33} \end{bmatrix} \cong \begin{bmatrix} 0.0119408 & 0.0050611 & -0.004504 \\ 0.0050611 & 0.016513 & -0.006219 \\ -0.004504 & -0.006219 & 0.0196583 \end{bmatrix}.$$

Al reemplazar [3.39], [3.40] y la inversa de \mathbf{R} en [3.41] y resolver el sistema de ecuaciones, se obtienen las siguientes soluciones:

$$\hat{\beta}_1 = 0.852; \quad \begin{bmatrix} \hat{a}_{oa} \\ \hat{a}_{ob} \\ \hat{a}_{oc} \\ \hat{a}_{o1} \\ \hat{a}_{o2} \\ \hat{a}_{o3} \\ \hat{a}_{o4} \\ \hat{a}_{o5} \\ \hat{a}_{o6} \\ \hat{a}_{o7} \\ \hat{a}_{o8} \\ \hat{a}_{o9} \\ \hat{a}_{o10} \end{bmatrix} = \begin{bmatrix} -0.168 \\ 0.281 \\ 0.908 \\ -0.254 \\ 0.110 \\ -0.492 \\ -0.060 \\ 0.212 \\ -0.525 \\ 0.264 \\ 0.128 \\ 0.332 \\ 0.809 \end{bmatrix}; \quad \begin{bmatrix} \hat{a}_{ma} \\ \hat{a}_{mb} \\ \hat{a}_{mc} \\ \hat{a}_{m1} \\ \hat{a}_{m2} \\ \hat{a}_{m3} \\ \hat{a}_{m4} \\ \hat{a}_{m5} \\ \hat{a}_{m6} \\ \hat{a}_{m7} \\ \hat{a}_{m8} \\ \hat{a}_{m9} \\ \hat{a}_{m10} \end{bmatrix} = \begin{bmatrix} -0.301 \\ -0.504 \\ -0.098 \\ -0.504 \\ -0.074 \\ -0.417 \\ 0.015 \\ -0.160 \\ -0.139 \\ -0.119 \\ 0.319 \\ 0.731 \\ 0.178 \end{bmatrix}; \quad \begin{bmatrix} \hat{a}_{na} \\ \hat{a}_{nb} \\ \hat{a}_{nc} \\ \hat{a}_{n1} \\ \hat{a}_{n2} \\ \hat{a}_{n3} \\ \hat{a}_{n4} \\ \hat{a}_{n5} \\ \hat{a}_{n6} \\ \hat{a}_{n7} \\ \hat{a}_{n8} \\ \hat{a}_{n9} \\ \hat{a}_{n10} \end{bmatrix} = \begin{bmatrix} -0.444 \\ 0.744 \\ -0.079 \\ -0.563 \\ 0.002 \\ -0.417 \\ 0.009 \\ -0.036 \\ -0.250 \\ 0.005 \\ 0.130 \\ 0.567 \\ 0.363 \end{bmatrix}; \quad \begin{bmatrix} \hat{e}_{m1} \\ \hat{e}_{m3} \\ \hat{e}_{m5} \\ \hat{e}_{m9} \end{bmatrix} = \begin{bmatrix} -0.181 \\ -0.248 \\ -0.067 \\ 0.511 \end{bmatrix} \quad [3.42]$$

Tal como ocurre con el MAM de R. L. Quaas, si se ajusta el modelo de abuela materna aquí presentado se reduce el número de animales a considerar en los vectores \mathbf{a}_o , \mathbf{a}_m y \mathbf{e}_m , y por ende en la matriz \mathbf{A} . Bajo este modelo, las incógnitas a resolver son:

$$\begin{aligned} \mathbf{a}_o' &= [a_{o1} \ a_{o2} \ a_{o3} \ a_{o4} \ a_{o5} \ a_{o6} \ a_{o7} \ a_{o8} \ a_{o9} \ a_{o10}] \\ \mathbf{a}_m' &= [a_{m1} \ a_{m2} \ a_{m3} \ a_{m4} \ a_{m5} \ a_{m6} \ a_{m7} \ a_{m8} \ a_{m9} \ a_{m10}] \\ \mathbf{a}_n' &= [a_{n1} \ a_{n2} \ a_{n3} \ a_{n4} \ a_{n5} \ a_{n6} \ a_{n7} \ a_{n8} \ a_{n9} \ a_{n10}] \\ \mathbf{e}_m' &= [e_{m1} \ e_{m3} \ e_{m5} \ e_{m9}] \end{aligned}$$

Nótese que en los vectores anteriores no se incluyó ninguna abuela materna fantasma porque en \mathbf{Z}_n se reemplazaron los valores de cría de abuela materna del animal 3 por $\frac{1}{2}a_{n1}$, de 7 por $-\frac{1}{3}a_{n2} + \frac{2}{3}a_{n5}$, y de 10 por $-\frac{1}{3}a_{n8} + \frac{2}{3}a_{n9}$. Las matrices de incidencia son iguales a:

$$\begin{aligned} \mathbf{Z}_o &= \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{Z}_m = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}; \\ \mathbf{Z}_n &= \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}; \quad \mathbf{Z}_p = \mathbf{I}_4 \end{aligned} \quad [3.43]$$

Aquí también $\mathbf{G} = \mathbf{A} \otimes \mathbf{G}_o$, pero \mathbf{A} es igual a:

$$A = \begin{bmatrix} 1 & 0 & 0.50 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.50 & 0 & 0.25 & 0 & 0 & 0.125 \\ 0.50 & 0 & 1 & 0 & 0 & 0.50 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0.50 & 0.50 & 0 & 0 & 0.25 \\ 0 & 0.50 & 0 & 0 & 1 & 0 & 0.50 & 0 & 0 & 0.25 \\ 0.25 & 0 & 0.50 & 0.50 & 0 & 1 & 0.25 & 0 & 0 & 0.125 \\ 0 & 0.25 & 0 & 0.50 & 0.50 & 0.25 & 1 & 0 & 0 & 0.50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.50 & 0.25 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & 1 & 0.50 \\ 0 & 0.125 & 0 & 0.25 & 0.25 & 0.125 & 0.50 & 0.25 & 0.50 & 1 \end{bmatrix} \quad [3.44]$$

Para los efectos ambientales maternos permanentes $\text{Var}(\mathbf{e}_m) = \mathbf{I}_4 \sigma_{Em}^2$. Finalmente:

$$R = \begin{bmatrix} 545 & 0 & 0 & 0 \\ 0 & 500 & 0 & 0 \\ 0 & 0 & 540 & 0 \\ 0 & 0 & 0 & 540 \end{bmatrix} \quad [3.45]$$

En la expresión [3.45], a la varianza del error (500) del individuo 1 se le sumó $\frac{3}{4}(60)$ en su elemento diagonal, mientras que a aquellos de los individuos 7 y 10 se adicionó $\frac{2}{3}(60)$ a su elemento diagonal. Al resolver [3.41] con [3.43], [3.44] y la inversa de [3.45] obtenemos:

$$\hat{\beta}_1 = 0.852; \quad \begin{bmatrix} \hat{a}_{o1} \\ \hat{a}_{o2} \\ \hat{a}_{o3} \\ \hat{a}_{o4} \\ \hat{a}_{o5} \\ \hat{a}_{o6} \\ \hat{a}_{o7} \\ \hat{a}_{o8} \\ \hat{a}_{o9} \\ \hat{a}_{o10} \end{bmatrix} = \begin{bmatrix} -0.254 \\ 0.110 \\ -0.492 \\ -0.060 \\ 0.212 \\ -0.525 \\ 0.264 \\ 0.128 \\ 0.332 \\ 0.809 \end{bmatrix}; \quad \begin{bmatrix} \hat{a}_{m1} \\ \hat{a}_{m2} \\ \hat{a}_{m3} \\ \hat{a}_{m4} \\ \hat{a}_{m5} \\ \hat{a}_{m6} \\ \hat{a}_{m7} \\ \hat{a}_{m8} \\ \hat{a}_{m9} \\ \hat{a}_{m10} \end{bmatrix} = \begin{bmatrix} -0.504 \\ -0.074 \\ -0.417 \\ 0.015 \\ -0.160 \\ -0.139 \\ -0.119 \\ 0.319 \\ 0.731 \\ 0.178 \end{bmatrix}; \quad \begin{bmatrix} \hat{a}_{n1} \\ \hat{a}_{n2} \\ \hat{a}_{n3} \\ \hat{a}_{n4} \\ \hat{a}_{n5} \\ \hat{a}_{n6} \\ \hat{a}_{n7} \\ \hat{a}_{n8} \\ \hat{a}_{n9} \\ \hat{a}_{n10} \end{bmatrix} = \begin{bmatrix} -0.563 \\ 0.002 \\ -0.417 \\ 0.009 \\ -0.036 \\ -0.250 \\ 0.005 \\ 0.130 \\ 0.567 \\ 0.363 \end{bmatrix}; \quad \begin{bmatrix} \hat{e}_{m1} \\ \hat{e}_{m3} \\ \hat{e}_{m5} \\ \hat{e}_{m9} \end{bmatrix} = \begin{bmatrix} -0.181 \\ -0.248 \\ -0.067 \\ 0.511 \end{bmatrix} \quad [3.46]$$

Como puede observarse, las soluciones [3.42] y [3.46] son iguales. Cabe agregar que la esperanza del vector de observaciones es la misma en ambos modelos, e igual a:

$$E(\mathbf{y}) = \mathbf{x} \hat{\beta}_1 = \begin{bmatrix} 180 \\ 161 \\ 151 \\ 162 \end{bmatrix} 0.852 = \begin{bmatrix} 153.36 \\ 137.17 \\ 128.65 \\ 138.02 \end{bmatrix} \quad [3.47]$$

Adicionalmente, la matriz de (co)varianzas se calculó a partir de la expresión siguiente:

$$V = Z_o A Z_o' 100 + Z_m A Z_m' 75 + Z_n A Z_n' 60 - 25(Z_o A Z_m' + Z_m A Z_o') + 15(Z_o A Z_n' + Z_n A Z_o') + 18(Z_m A Z_n' + Z_n A Z_m') + Z_p Z_p' 50 + R \quad [3.48]$$

Y para ambos modelos fue igual a

$$V = \begin{bmatrix} 785.50 & 118.125 & 0 & 0 \\ 118.125 & 785.50 & 25 & 12.50 \\ 0 & 25 & 785.50 & 45.625 \\ 0 & 12.50 & 45.625 & 785.50 \end{bmatrix} \quad [3.49]$$

3.4. DISCUSIÓN

Para especificar correctamente la fracción aditiva de las varianzas y las covarianzas cuando la abuela materna es desconocida, Van Vleck (1990a) sugirió incluir las abuelas no identificadas dentro del vector de valores de cría, asumiéndolas no emparentadas entre sí, y con un solo nieto cada una. Sin embargo, dado que se deben incluir todas las abuelas maternas no identificadas o “fantasmas” en el archivo de datos, en poblaciones con una alta fracción de abuelas desconocidas (razas de formación reciente o pedigrí abierto) se produce un aumento importante en el número de ecuaciones a resolver: por cada abuela fantasma que se agrega en la matriz A , se aumenta el vector de soluciones en 3 ecuaciones. En esta investigación se presenta un enfoque alternativo al desarrollar una solución que no requiere agrandar el orden de las ecuaciones a resolver. La solución consiste en expresar el valor de cría de abuela materna de una hembra desconocida cuya hija es madre de un animal con registro, como una función lineal del valor de cría de su hija y del padre de su nieto (ecuación [3.25]). Si el abuelo materno del nieto es también desconocido, entonces su valor de cría de abuela materna es reemplazado por la mitad del valor de cría de abuela materna de su hija (ecuación [3.26]). Nótese que esta idea sugerida por R. L. Quaas para los efectos maternos, evita tener que agregar en los vectores de valores de cría aquellas vacas “fantasmas”. Tal como se mostró a lo largo del capítulo, para que el modelo aquí propuesto resulte equivalente al modelo de Willham (1972), es necesario agregar a la varianza del error la fracción de varianza aditiva que no es contemplada por el modelo de evaluación. En las ecuaciones [3.27] y [3.28] se muestran dichas fracciones en función que el abuelo materno sea conocido o desconocido, respectivamente. La única modificación a estas expresiones ocurre cuando el abuelo materno es consanguíneo. En tal situación, en vez de utilizarse [3.27] y [3.28] deben emplearse [3.29] y [3.30]. Consecuentemente, esta propuesta no sólo permite disminuir marcadamente el número de ecuaciones a resolver sino que además, la matriz R conserva su estructura diagonal y no constituye un obstáculo, ni para el cálculo de la matriz de los coeficientes de las ecuaciones de modelo mixto, ni para la estimación de los componentes de (co)varianza.

Capítulo 4

Predicción de los valores de cría de animales nacidos por trasplante embrionario, en el modelo animal con efectos maternos.

4.1. INTRODUCCIÓN

La aplicación de los esquemas de ovulación múltiple y transferencia embrionaria (MOET) en los programas de mejoramiento genético en bovinos de carne requieren la evaluación temprana del mérito genético de grupos de hermanos enteros para reducir el intervalo generacional y, en consecuencia, incrementar el progreso genético (Smith, 1984). Sin embargo, mientras que en los programas de evaluación genética clásica, la madre del ternero es la que aporta los efectos genéticos directos y los maternos, en los esquemas MOET existen distintas madres receptoras que proveen el ambiente materno a las crías de una misma madre biológica. Estas vacas receptoras pueden ser vacas de razas lecheras que están finalizando su vida productiva, vacas cruce comerciales, o pueden ser vacas de pedigrí que pueden estar emparentadas con las madres biológicas. En consecuencia, los modelos de evaluación genética deben ser modificados para que los animales nacidos por transferencia embrionaria contemplen “tres padres”: sus dos padres biológicos y la vaca receptora (Van Vleck, 1990b). Para ello, el sistema de registros requiere que se tenga identificada la raza y la edad de la receptora, además de la madre y el padre del individuo (Schaeffer y Kennedy, 1989). En muchas situaciones de campo, no se cuenta con la identificación de la vaca receptora: 7.25% de los animales de la evaluación genética del Brangus argentino, y 3.87% de los animales en la evaluación Braford argentina provienen de terneros nacidos por transferencia embrionaria cuyas madres receptoras no están identificadas. Ante esta situación, se presentan dos problemas. Por un lado, el desconocimiento del valor apropiado de la edad de la madre (EM) sesga la predicción del valor de cría del ternero. En adición, se genera una covarianza residual entre los datos de animales con la misma madre receptora, la cual es generalmente utilizada en más de una ocasión. Esto produce una incorrecta especificación de la varianza residual y lleva a que los predictores no sean de mínima varianza. Para resolver el primer problema, en el presente capítulo se propone utilizar una metodología estadística usada en epidemiología que permite atenuar el sesgo producido por error de medición clásico en el efecto fijo EM. La justificación de su uso se debe a que la evaluación genética animal empleando modelos mixtos y predicción lineal insesgada de mínima varianza (BLUP, Henderson, 1984) se realiza condicional a que las variables consideradas como efectos fijos (sexo, edad del animal, edad de la madre, grupo de contemporáneos, etc), o aleatorios (los valores de cría, los efectos ambientales maternos permanentes), sean medidas sin error. En la práctica la variable explicativa EM es observada en una fracción mayoritaria de los registros pero no en todos. El hecho que las EM desconocidas no sean datos perdidos al azar (*‘missing at random’*) dado que ocurre parcialmente en un grupo de animales de alto valor genético, invalida el uso de metodologías de “datos faltantes” (Im et al., 1989). Adicionalmente, para contemplar la covarianza entre los residuos de animales nacidos por trasplante embrionario con la misma madre receptora desconocida, se propone el uso de técnicas de suavizado de los residuales siguiendo los lineamientos de

Eilers y Marx (1996) con *splines* penalizadas (*P-splines*), mediante el enfoque del modelo mixto de Cantet et al. (2005).

4.2. METODOLOGÍA PARA ATENUAR EL SESGO POR EDAD DE MADRE INCIERTA

En esta sección se presenta una metodología utilizada en estudios epidemiológicos que permite atenuar el sesgo producido por error de medición clásico, en el efecto fijo EM. Se mostrará cómo corregir las ecuaciones de modelo mixto cuando se presentan errores de medición aleatorios, es decir, en cualquier categoría del efecto fijo, y se conoce el factor de atenuación λ (ver definición en la sección 4.2.1.1). Finalmente, la metodología se aplicará para analizar la edad de la madre en el peso al nacer y al destete de vacunos Brangus en la Argentina.

4.2.1. Modelos con error de medición.

Para abordar esta sección, se seguirá la notación tradicional de la literatura de modelos con error de medición (MEM, Buzas et al., 2005). Llamaremos Y a la variable respuesta, Z a una covariable medida sin error, X a la covariable no observada en todos los datos, W a la covariable observada que actúa como variable sucedánea (*surrogate*) de X y e es el error aleatorio, el cual es independiente de X , Z y W . El primer paso consiste en identificar la estructura del error, es decir, conocer como la variable W (observada) se relaciona con los valores verdaderos X . Si bien existen numerosas relaciones entre ambas variables, en la práctica se utilizan sólo dos tipos (Carroll et al., 1995):

- Modelos de error: incluye los modelos clásicos de error de medición y los modelos de calibración. Se modela la distribución condicional de W dado (Z, X) . Su uso es adecuado cuando se trata de determinar X directamente, aunque su observación se vea dificultada por diversas causas que generan errores de medición.
- Modelos de calibración de la regresión: incluye los modelos de variable controlada o modelos de error de Berkson. Se modela la distribución condicional de X dado (Z, W) . Se aplica especialmente en estudios de laboratorio (Berkson, 1950).

En el caso de la EM desconocida, el modelo más apropiado es el modelo de error clásico: se intenta conocer la edad de la madre de manera precisa (X) pero, a causa de errores de medición – por ejemplo en la toma de datos – sólo se observa la edad de madre desconocida (W). A continuación se describirá la especificación del error de medición “clásico” (Buzas et al., 2005), y se analizarán sus consecuencias sobre el sesgo y la varianza residual. Se asume normalidad multivariada de la distribución conjunta $f(Y, Z, X, W)$ y de todas las marginales.

4.2.1.1. Modelo clásico de error de medición

El error de medición clásico es igual a (Fuller, 1987; Buzas et al., 2005):

$$w_i = x_i + u_i \quad [4.1]$$

donde u_i es el error de medición tal que $u_i \sim N(0, \sigma_u^2)$, e independiente de x_i . A su vez, $x_i \sim N(\mu_x, \sigma_x^2)$. En la literatura referida a los MEM se dice que el error de medición es ‘no diferencial’ (aleatorio) cuando las siguientes distribuciones conjuntas son iguales:

$$f(Y|Z, X, W) = f(Y|Z, X) \quad [4.2]$$

Dicho de otro modo: el error es no diferencial cuando la variable W no genera información adicional en el caso que X fuese conocida e incluida en el modelo. Esto sucede, por ejemplo, con EM, o con la proporción de genes de una raza dentro de la evaluación genética en bovinos de carne, en donde el error de medición puede darse en cualquiera de las categorías del efecto fijo, y no necesariamente en una en particular. Adicionalmente, bajo este modelo, la variable sucedánea w es insesgada con respecto a x :

$$\begin{aligned} E(w|x) &= E(w) + \frac{\text{cov}(w, x)}{\text{Var}(x)} (x - E(x)) \\ &= E(x + u) + \frac{\text{cov}(x + u, x)}{\sigma_x^2} (x - \mu_x) \\ &= \mu_x + \frac{\sigma_x^2}{\sigma_x^2} (x - \mu_x) \\ &= \mu_x + x - \mu_x \\ &= x \end{aligned} \quad [4.3]$$

Previo a desarrollar un modelo animal con errores de medición, se presentará la metodología de error de medición clásico para un modelo de regresión múltiple. Tomando como parámetros a β_Z y β_X , suponga que las observaciones de los registros $i = 1, \dots, n$, siguen el siguiente modelo verdadero:

$$y_i = z_i \beta_Z + x_i \beta_X + e_i \quad [4.4]$$

donde $e_i \sim N(0, \sigma_e^2)$, e independiente de x_i . A su vez las variables z_i y x_i no están correlacionadas. Bajo el supuesto de error no diferencial:

$$\text{Cov}(e_i, w_i) = 0 \quad [4.5]$$

Por lo tanto,

$$\begin{aligned}
 \text{Cov}(y_i, w_i) &= \text{cov}[z_i \beta_z + x_i \beta_x + e_i, w_i] \\
 &= \text{cov}[z_i \beta_z + x_i \beta_x + e_i, x_i + u_i] \\
 &= \text{cov}[x_i \beta_x, x_i] \\
 &= \beta_x \text{cov}[x_i, x_i] \\
 &= \beta_x \sigma_x^2
 \end{aligned} \tag{4.6}$$

Sin embargo, el modelo [4.4] no puede ser ajustado dado que x_i no es observable. En su lugar, se mide la variable sucedánea w_i :

$$y_i = z_i \beta_z^* + w_i \beta_x^* + \varepsilon_i \tag{4.7}$$

donde $E(\varepsilon_i) = 0$. El objetivo es estimar los parámetros bajo este modelo para, posteriormente, compararlos con los parámetros del modelo verdadero presentado en [4.4] y, finalmente, estudiar la relación entre ambos estimadores. Para ello se definirá al factor de atenuación (o ‘*Reliability Ratio*’) del siguiente modo:

$$\lambda = \frac{\text{Var}(x_i)}{\text{Var}(w_i)} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \tag{4.8}$$

y a su inversa se la denominará corrección lineal para la atenuación (‘*linear correction for attenuation*’). La esperanza del estimador $\hat{\beta}_x^*$ es igual a:

$$E(\hat{\beta}_x^*) = \frac{\text{cov}[y_i, w_i]}{\text{cov}[w_i, w_i]} = \frac{\beta_x \sigma_x^2}{\text{Var}(w_i)} = \frac{\beta_x \sigma_x^2}{\text{Var}(x_i + u_i)} = \frac{\beta_x \sigma_x^2}{\sigma_x^2 + \sigma_u^2} = \beta_x \lambda \tag{4.9}$$

Tal como puede observarse en la ecuación [4.9], el error de medición tiende a sesgar al coeficiente β_x hacia cero. Este efecto es normalmente conocido como atenuación o atenuación hacia el nulo (‘*attenuation or attenuation to the null*’). Con respecto al estimador $\hat{\beta}_z^*$, su esperanza es igual a (Carroll, 2005, pág. 15):

$$E(\hat{\beta}_z^*) = \beta_z + \beta_x (1 - \lambda) \gamma_z \tag{4.10}$$

donde $\gamma_z = \beta_z \frac{\text{cov}(x_i, z_i)}{\text{Var}(z_i)}$. Sin embargo, cuando la variable z_i es independiente de x_i ,

$\gamma_z = 0$. En consecuencia, el coeficiente β_z es insesgado:

$$E(\hat{\beta}_z^*) = \beta_z \tag{4.11}$$

Por último, al regresar en la variable w , la varianza de y_i resulta igual a:

$$\begin{aligned}
 \text{var}(y_i | w_i) &= \text{var}(y_i) (1 - \rho_{yw}^2) = \sigma_y^2 \left(1 - \frac{\text{cov}(y_i, w_i)^2}{\text{var}(y_i) \text{var}(w_i)} \right) = \sigma_y^2 \left(1 - \frac{\beta_x^2 (\sigma_x^2)^2}{\sigma_y^2 \text{var}(w_i)} \right) \\
 &= \sigma_y^2 - \sigma_y^2 \frac{\beta_x^2 (\sigma_x^2)^2}{\sigma_y^2 \text{var}(w_i)} = \sigma_e^2 + \beta_x^2 \sigma_x^2 \left(1 - \frac{\sigma_x^2}{\text{var}(w_i)} \right) = \sigma_e^2 + \beta_x^2 \sigma_x^2 \left(1 - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \right) \\
 &= \sigma_e^2 + \beta_x^2 \sigma_x^2 \left(\frac{\sigma_x^2 + \sigma_u^2 - \sigma_x^2}{\sigma_x^2 + \sigma_u^2} \right) = \sigma_e^2 + \beta_x^2 \sigma_x^2 \left(\frac{\sigma_u^2}{\sigma_x^2 + \sigma_u^2} \right) = \sigma_e^2 + \beta_x^2 \sigma_u^2 \left(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \right) \\
 &= \sigma_e^2 + \beta_x^2 \sigma_u^2 \lambda
 \end{aligned}
 \tag{4.12}$$

Nótese que la varianza de una observación con error de medición es superior, en una magnitud de $\beta_x^2 \sigma_u^2 \lambda$, a una observación donde X es medida sin error. Este resultado afirma el hecho que siempre resulta menos informativo observar W que observar X .

4.2.1.2. Corrección por sesgo

Tal como se mostró en el apartado anterior, bajo el modelo con error de medición clásico, el estimador de β_x es sesgado (Carroll 2005; Buzas et al., 2005). Sin embargo, cuando el factor de atenuación es conocido se puede aplicar una técnica sencilla denominada corrección por atenuación (*'correction for attenuation'*) para eliminar el sesgo. Esta metodología permite obtener estimaciones insesgadas de β_x simplemente dividiendo a la solución del nivel del efecto fijo medido con error ($\hat{\beta}_x^*$) por λ , y ajustando por la varianza adicional que surge al no utilizarse el modelo verdadero. La expresión del coeficiente corregido por sesgo es:

$$\hat{\beta}_x = \frac{\hat{\beta}_x^*}{\hat{\lambda}}
 \tag{4.13}$$

El factor de atenuación se debe estimar a partir de los datos: $\hat{\sigma}_x^2$ se estima mediante los registros de los individuos cuya EM es conocida y, $\hat{\sigma}_u^2$ a través de los animales con EM desconocida. Una vez obtenidas ambas varianzas, $\hat{\lambda}$ se calcula tal como se muestra en [4.8]. En el caso del Brangus argentino, el 92.75% de los datos se utiliza para calcular $\hat{\sigma}_x^2$ mientras que el 7.25% restante para obtener $\hat{\sigma}_u^2$. Estos porcentajes cambian al 96.13 % y 3.87 % en pesos al destete de la raza Braford argentina.

4.2.1.3. Modelo animal con error de medición clásico en el efecto fijo edad de madre

La ecuación básica de un modelo animal con efectos maternos (Quaas y Pollak, 1980), para el dato del individuo i , hijo de la madre j , al cual se le incluye error de medición en el efecto fijo EM es igual a:

$$y_{ij} = \mathbf{x}_{1i} \boldsymbol{\beta}_1 + w_i \beta_2 + a_{oi} + a_{mj} + e_{mj} + e_{oi} \quad [4.14]$$

donde \mathbf{x}_{1i} es la i -ésima fila de la matriz de incidencia de los efectos fijos conocidos contenidos en el vector $\boldsymbol{\beta}_1$. El escalar β_2 está asociado con el efecto fijo EM desconocida. La variable indicadora w_i , del tipo “dummy” (Searle, 1971, cáp. 4), es la que garantiza que el segundo término de [4.14] sólo esté presente en los datos de animales con EM incierta (toma el valor de 1 si la EM es desconocida ó 0 en caso contrario). Los valores aleatorios a_{oi} y a_{mj} son el valor de cría directo del individuo i y el valor de cría materno de j (madre de i), respectivamente; e_{mj} es el efecto ambiental materno permanente de la madre j . Por último, la variable aleatoria e_{oi} es el error del modelo.

Para expresar [4.14] en notación matricial, sea \mathbf{y} el vector de observaciones tal que $\mathbf{y} = [y_{ij}]$. Este vector es particionado en $\mathbf{y}' = [\mathbf{y}_c', \mathbf{y}_d']'$. La notación diferencia los datos de individuos con información: 1) completa (C); 2) incierta (D). El orden de \mathbf{y}_c es $n_C \times 1$ y el de \mathbf{y}_d es $n_D \times 1$. Además, $n = n_C + n_D$. Esto permite escribir al modelo animal unicarácter con efectos maternos y error de medición clásico en el efecto fijo EM (MEM) del modo siguiente:

$$\begin{bmatrix} \mathbf{y}_c \\ \mathbf{y}_d \end{bmatrix} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{w} \beta_2 + \mathbf{Z}_o \mathbf{a}_o + \mathbf{Z}_m \mathbf{a}_m + \mathbf{Z}_p \mathbf{e}_m + \begin{bmatrix} \mathbf{e}_{oc} \\ \mathbf{e}_{od} \end{bmatrix} \quad [4.15]$$

donde \mathbf{X}_1 es la matriz ($n \times p_1$) de incidencia de los efectos fijos medidos sin error; $\boldsymbol{\beta}_1$ es un vector ($p_1 \times 1$) de efectos fijos conocidos. El vector \mathbf{w} posee elementos iguales a 0 ó 1, y se lo puede representar del siguiente modo:

$$\mathbf{w} = \begin{bmatrix} \mathbf{0}_{nC \times 1} \\ \mathbf{1}_{nD \times 1} \end{bmatrix} \quad [4.16]$$

Las matrices de incidencia \mathbf{Z}_o y \mathbf{Z}_m , relacionan los valores de cría directos (\mathbf{a}_o) y maternos (\mathbf{a}_m) con las observaciones, mientras que \mathbf{Z}_p asocia los datos a los efectos maternos permanentes (\mathbf{e}_m). Por último, los errores aleatorios del modelo están representados por el vector particionado $[\mathbf{e}_{oc} \ \mathbf{e}_{od}]'$. En ausencia de selección, todos los vectores aleatorios a la derecha de [4.15] tienen esperanza $\mathbf{0}$, en tanto que la estructura de varianzas y covarianzas es igual a:

$$\text{Var} \begin{bmatrix} \mathbf{a}_o \\ \mathbf{a}_m \\ \mathbf{e}_m \\ \mathbf{e}_o \end{bmatrix} = \begin{bmatrix} A\sigma_{Ao}^2 & A\sigma_{AoAm} & \mathbf{0} & \mathbf{0} \\ A\sigma_{AoAm} & A\sigma_{Am}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_d \sigma_{Em}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \quad [4.17]$$

En [4.17], la matriz A contiene las relaciones aditivas (Henderson, 1984); σ_{Ao}^2 es la varianza aditiva por efectos directos y σ_{Am}^2 por efectos maternos, mientras que σ_{AoAm} es la covarianza aditiva entre ambos efectos. La varianza de los errores maternos permanentes es σ_{Em}^2 y, σ_e^2 es la varianza de los errores. Finalmente, la matriz R es igual a:

$$R = \begin{bmatrix} I_{n_c} \sigma_e^2 & 0 \\ 0 & I_{n_D} [\beta_2^2 \sigma_u^2 \lambda + \sigma_e^2] \end{bmatrix} \quad [4.18]$$

Cabe destacar que la matriz diagonal representada en [4.18] es la que se utiliza para armar las ecuaciones de modelo mixto. En el caso del escalar β_2 , se utiliza la estimación del año anterior por ser su valor equivalente al promedio de las soluciones de los últimos años. Esto sucede porque las soluciones del efecto fijo EM son muy similares entre años. Por último, las ecuaciones de modelo mixto corregidas por sesgo son las siguientes:

$$\begin{bmatrix} X_1' R^{-1} X_1 & X_1' R^{-1} w \lambda^{-0.5} & X_1' R^{-1} Z_o & X_1' R^{-1} Z_m & X_1' R^{-1} Z_p \\ \lambda^{-0.5} w' R^{-1} X_1 & \lambda^{-1} w' R^{-1} w & \lambda^{-0.5} w' R^{-1} Z_o & \lambda^{-0.5} w' R^{-1} Z_m & \lambda^{-0.5} w' R^{-1} Z_p \\ Z_o' R^{-1} X_1 & Z_o' R^{-1} w \lambda^{-0.5} & Z_o' R^{-1} Z_o + A^{-1} g^{11} & Z_o' R^{-1} Z_m + A^{-1} g^{12} & Z_o' R^{-1} Z_p \\ Z_m' R^{-1} X_1 & Z_m' R^{-1} w \lambda^{-0.5} & Z_m' R^{-1} Z_o + A^{-1} g^{21} & Z_m' R^{-1} Z_m + A^{-1} g^{22} & Z_m' R^{-1} Z_p \\ Z_p' R^{-1} X_1 & Z_p' R^{-1} w \lambda^{-0.5} & Z_p' R^{-1} Z_o & Z_p' R^{-1} Z_m & Z_p' R^{-1} Z_p + I_d \frac{1}{\sigma_{Em}^2} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{a}_o \\ \hat{a}_m \\ \hat{e}_m \end{bmatrix} = \begin{bmatrix} X_1' R^{-1} y \\ \lambda^{-0.5} w' R^{-1} y \\ Z_o' R^{-1} y \\ Z_m' R^{-1} y \\ Z_p' R^{-1} y \end{bmatrix} \quad [4.19]$$

$$\text{donde } \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix} = \begin{bmatrix} \sigma_{Ao}^2 & \sigma_{AoAm} \\ \sigma_{AoAm} & \sigma_{Am}^2 \end{bmatrix}^{-1}$$

4.2.1.4. Efecto de la metodología MEM sobre la EM incierta en Brangus argentino

Se analizaron 137,304 registros de animales nacidos entre 1974 y 2010 pertenecientes al programa ERBra de la Asociación Argentina de Brangus. En el cuadro 4.1 se presentan las soluciones para los efectos de EM para pesos al nacer (PN) y al destete (PD) con dos modelos de análisis distintos: Convencional y MEM. El modelo “Convencional” es un modelo animal multicarácter que incluye los pesos al nacer, al destete (componente directa y materna) y el peso ‘final’ (18-24 meses de edad) con componentes de (co)varianza obtenidos previamente sobre los mismos datos. En el MEM se modela el efecto de la variable medida con error (W) tal como aparece en [4.1], siendo la ecuación básica del modelo igual a la expresión [4.14]. Asimismo, la varianza residual del carácter para los animales con EM incierta fue aumentada, desde σ_e^2

(estimada oportunamente como 478 kg²) a $\beta_x^2 \sigma_x^2 \left(\frac{\sigma_u^2}{\sigma_x^2 + \sigma_u^2} \right) + \sigma_e^2$ (ecuación [4.12]). La

varianza σ_x^2 fue estimada por Cantet (2005) a partir de un modelo con todos los efectos fijos del modelo de evaluación genética salvo la edad de la madre que fue considerada aleatoria. La varianza de este efecto fue diagonal, con elementos iguales a σ_x^2 para los

animales donde la EM era conocida con exactitud, y a σ_u^2 cuando al edad de la madre fue incierta. El modelo fue ajustado con PROC MIXED de SAS y la opción de covarianza UN(1) para el efecto aleatorio. Hubo 96470 animales con EM exacta, y 40834 registros donde la EM era incierta. Los valores estimados fueron los siguientes: $\sigma_x^2 = 0.724\sigma_e^2$ y $\sigma_u^2 = 0.155\sigma_e^2$. En consecuencia, la inversa del cociente $\sigma_x^2 / (\sigma_x^2 + \sigma_u^2)$ fue igual a 1.214. En ambos modelos, se agruparon en la categoría “EM desconocida” a los animales con EM incierta, ya sean nacidos por transplante embrionario o con edades de madre desconocida.

Cuadro 4.1: Soluciones para el efecto EM para peso al nacer y al destete con dos modelos en bovinos de carne de la raza Brangus.

	Convencional		MEM	
EM	PN	PD	PN	PD
2	−0.882	1.837	−1.105	1.787
3	−0.552	3.839	−0.775	3.789
4	−0.003	8.012	−0.225	7.962
5-9	0.319	11.642	0.097	11.592
10 ó más	−0.159	10.318	−0.381	10.267
Desconocida	0.105	9.545	−0.102	7.815

En el Cuadro 4.1 se observa que el descuento que sufren los animales con EM Desconocida es mayor en el modelo Convencional que en el MEM, tanto para el PN (0.105 vs. −0.102) como para el PD (9.545 vs. 7.815).

4.3. METODOLOGÍA PROPUESTA PARA CONSIDERAR LA COVARIANZA RESIDUAL

4.3.1. Suavización de covarianza residual en rodeos con datos de transplantes embrionarios

Cuando se presentan datos de animales nacidos por transplante embrionario cuyas madres receptoras no se encuentran identificadas, es difícil justificar que los errores del modelo son independientes. Esto se debe a que las vacas receptoras suelen ser utilizadas en más de una ocasión dentro del mismo rodeo, generando covarianzas entre los residuos de los animales en evaluación. Sin embargo, la falta de información genealógica completa en los archivos de datos no permite identificar la frecuencia con la cual se utiliza la misma vaca receptora. Como consecuencia de ello, se desconoce la forma funcional en la que se relacionan los residuales de los animales en el tiempo: su estructura de covarianzas. Los modelos semiparamétricos son una alternativa viable que permiten relajar el supuesto restrictivo acerca de la forma funcional conocida. En la

presente investigación, se consideraron modelos semiparamétricos para los errores en los que el suavizado de los residuales se realizó mediante el ajuste de *splines* penalizados (*P-splines*, Eilers y Marx, 1996) sobre la covariable tiempo (representada aquí por el día de nacimiento – DN – del animal, expresado en días julianos), a través del enfoque de modelo mixto desarrollado por Cantet y colaboradores (2005). La metodología se aplicó únicamente a los establecimientos con animales nacidos por trasplante embrionario. A tal efecto, utilizaremos un enfoque metodológico similar al de Wade et al. (1993), quienes ajustaron un proceso autoregresivo de orden uno, AR(1), al efecto de grupo de contemporáneos, el cual fue considerado como una serie de tiempo (Box y Jenkins, 1976). En este mismo sentido, la estrategia de Box y Jenkins (1976), se aplicó a cada una de las series de tiempo bajo estudio. La principal ventaja del ajuste de procesos AR(1) radica en que Wade y Quaas (1993) desarrollaron un procedimiento que facilita la inversión de la matriz de covarianzas resultante de modo de poder escribir las ecuaciones de modelo mixto sin dificultad.

4.3.2. Observaciones

Se emplearon los residuales del carácter peso al destete de rodeos con más de 80 animales nacidos por trasplante embrionario y que contaban con todas las categorías del efecto fijo edad de la madre. Los datos provienen del programa de evaluación genética ERBra, Asociación Argentina de Brangus. Sólo se utilizaron series de años consecutivos e ininterrumpidos. Una vez elegidos los 12 establecimientos, se calcularon los residuales mediante el uso de las soluciones obtenidas en la corrida del ERBra del año 2011. Se estudió la normalidad de los residuales mediante el gráfico de probabilidad normal (*Normal Probability Plot*), y el test de Shapiro-Wilk (Shapiro y Wilk, 1965), usando el procedimiento PROC Univariate de SAS (SAS Institute Inc, 2012). En el presente análisis no fue necesario transformar los datos: en ninguno de los establecimientos se encontraron evidencias de no-normalidad de los residuales. Para determinar el largo de la serie, se tomó el valor 1 para el día juliano coincidente con la fecha de nacimiento del primer ternero del rodeo. En el Cuadro 4.2 se detalla el número de observaciones, el número y el porcentaje de trasplantes embrionarios, los años de registros y el largo de la serie de cada uno de los establecimientos bajo estudio.

Cuadro 4.2: Descripción de los establecimientos de la raza Brangus.

Establecimiento	Nº de observaciones	Nº de TE	Años de registros	Largo de la serie [Días Julianos]
I	2993	1343 (44.87 %)	1996 - 2010	5317
II	1425	947 (66.46 %)	1999 - 2010	4112
III	3046	753 (24.72 %)	2001 - 2010	3384
IV	1740	459 (26.38 %)	2002 - 2010	2890
V	15087	445 (2.95 %)	1981 - 2009	10547
VI	2106	354 (16.81 %)	2003 - 2010	2644
VII	509	317 (62.28 %)	2003 - 2010	2166
VIII	1371	182 (13.27 %)	2003 - 2010	2778
IX	4568	151 (3.30 %)	1997 - 2010	4892
X	2540	172 (6.77 %)	2002 - 2010	3015
XI	2907	84 (2.89 %)	2001 - 2010	3136
XII	1297	144 (11.10 %)	2004 - 2010	2124

En el Cuadro 4.2 se evidencia la variabilidad existente entre establecimientos, en lo que respecta al número de observaciones, número y porcentaje de transplantes embrionarios, años de registros y largo de las series. De hecho, no fue posible agrupar las cabañas bajo ninguna de estas categorías. Asimismo, dado que en los establecimientos los registros de los terneros no están equiespaciados, y este supuesto es fundamental en los análisis de series de tiempo, se suavizaron los residuales mediante *P-splines* cúbicos, cuya teoría se desarrolla en la sección siguiente.

4.3.3. *Splines penalizadas*

La palabra inglesa *spline* define una regla flexible sujeta a dos puntos, utilizada en dibujo técnico para pintar curvas suaves pasando por puntos especificados. Estas condiciones de suavidad y adaptabilidad son las que hacen que las funciones polinomiales a trozos adopten este nombre. Una función *spline* es un polinomio a trozos de grado definido y con ciertas propiedades de regularidad (derivabilidad). En general, las *B-splines* de grado k (Eilers y Marx, 1996):

- Consisten en $k + 1$ segmentos polinómicos, cada uno de grado k , conectados entre sí;
- Las piezas polinómicas se unen en k nodos;

- En los puntos de unión (nodos), las derivadas de orden $k - 1$ son continuas;
- Una *B-spline* de grado d es positiva en el dominio expandido por $k + 2$ nodos, y cero en cualquier otro punto;
- Se superponen en $2k$ piezas polinómicas con los nodos vecinos, con excepción de los extremos;
- En cualquier punto del dominio de la variable, $k + 1$ *B-splines* son distintas a cero;

En cuanto al número de nodos, autores como Ruppert et al. (2003) aconsejan elegir los K nodos en los K -cuantiles de la variable x , es decir, cada nodo t_k sería el cuantil $k/(K+1)$ de x . Sin embargo, la mayoría utilizan como regla:

$$\text{Número de nodos} = \text{mín. } \{40; \text{valores únicos de } x/4\}$$

4.3.3.1. *B-splines utilizando la covariable día de nacimiento*

Para abordar esta sección, se seguirá estrechamente los trabajos de Eilers y Marx (1996) y Cantet et al. (2005). Supongamos que contamos con residuales ($e_{i(l)}^*$) y variables explicatorias que dependen funcionalmente del tiempo (t_i). Entonces, un modelo de suavizado para residuos normales pertenecientes al rodeo l viene dado por la expresión:

$$e_{i(l)}^* = f(t_i)_{(l)} + e_{i(l)}, \quad e_{i(l)} \sim N(0, \sigma_{Eo}^2) \quad [4.20]$$

donde el tiempo t_i es el día de nacimiento (DN) del ternero, expresado en días julianos. A su vez, $f(t_i)_{(l)}$ es una función suave de los residuales (a la que no se le impone ninguna forma), la cual se puede escribir como:

$$f(t_i)_{(l)} = \sum_{i=1}^{nx} \mathbf{B}_{i(l)}^{(k)} \mathbf{b}_{i(l)} \quad [4.21]$$

En [4.21], la función $f(t_i)_{(l)}$ es una combinación lineal del vector de coeficientes

splines $\mathbf{b}_{i(l)} = [b_1, b_2, \dots, b_{nx}]'$ (los cuales deben ser estimados), y de

$\mathbf{B}_{i(l)} = [\mathbf{B}_1^{(k)}, \mathbf{B}_2^{(k)}, \dots, \mathbf{B}_{nx}^{(k)}]'$, que es un vector columna cuyos elementos son las funciones básicas *B-splines* (De Boor, 1993). Las funciones *splines* cúbicas ($k = 3$) son las más utilizadas en la práctica porque proveen curvas suaves, son fáciles de evaluar, se pueden incorporar en las ecuaciones de modelo mixto y, generalmente, arrojan resultados satisfactorios (Reinsch, 1967; Verbyla et al., 1999; White et al., 1999). Para cubrir el rango de la covariable DN, se necesitan $nx + 3$ funciones básicas $\mathbf{B}_i^{(k)}$, es decir, se deben adicionar 3 nodos en un extremo (Eilers y Marx, 1996, página 90). Cada una de estas funciones base puede calcularse empleando las formulas recursivas de Cox-De Boor (De Boor, 1972; De Boor, 1993), las cuales funcionan aún cuando los nodos no son equidistantes.

4.3.3.2. Algoritmo de Cox-De Boor

Sean $t_1, t_2, \dots, t_{nx+3}$ el conjunto de nodos distribuidos a lo largo del rango de la covariable DN. El algoritmo consiste en calcular $\mathbf{B}_i^{(0)}$ el cual es igual a 1 si DN está incluido en el intervalo $[t_i, t_{i+1})$, ó es 0 en caso de no estarlo. Por ejemplo, si $t_{14} = 4501$ y $t_{15} = 4968$, entonces para $DN = 4627$, $\mathbf{B}_{14}^{(0)} = 1$ mientras que $\mathbf{B}_1^{(0)}, \dots, \mathbf{B}_{13}^{(0)}, \mathbf{B}_{15}^{(0)}, \dots$, y $\mathbf{B}_{nx+3}^{(0)}$ son todos iguales a 0. Luego, se procede a obtener los coeficientes correspondientes a los $\mathbf{B}_i^{(1)}$ s, los $\mathbf{B}_i^{(2)}$ s, y por último, los $\mathbf{B}_i^{(3)}$ s, siguiendo el esquema representado en la Figura 4.1, llamado “*Blossom*” en donde se exponen las relaciones de dependencia entre las funciones de base de distintos órdenes.

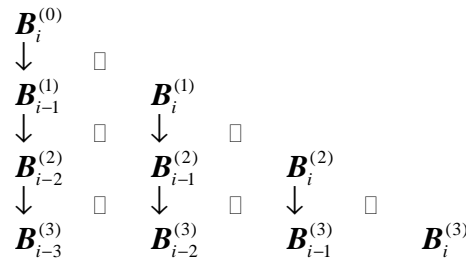


Figura 4.1: Representación esquemática de las relaciones funcionales entre funciones básicas *B-splines* de órdenes crecientes

Tal como se observa en la Figura 4.1, para cualquier orden $k > 1$, cada una de las funciones básicas $\mathbf{B}_i^{(k)}$ se obtiene como una combinación lineal de dos funciones básicas de orden $k - 1$. A su vez, las funciones base de orden k cumplen la condición de sumar a 1, es decir $\sum_{i=1}^{nx} \mathbf{B}_i^{(k)} \equiv 1$. Véase también que, para expresar $\mathbf{B}_i^{(3)}$, se necesita conocer $\mathbf{B}_{i-3}^{(3)}$, $\mathbf{B}_{i-2}^{(3)}$, $\mathbf{B}_{i-1}^{(3)}$ y $\mathbf{B}_i^{(3)}$. A continuación se muestran las fórmulas recursivas de Cox-De Boor:

$$\mathbf{B}_i^{(1)} = \begin{cases} 1 & \text{si } DN \in [t_i, t_{i+1}) \\ 0 & \text{en otro lado} \end{cases} \quad [4.22]$$

y

$$\mathbf{B}_i^{(m)} = \frac{DN - t_i}{t_{i+m-1} - t_i} \mathbf{B}_i^{(m-1)} + \frac{t_{i+m} - DN}{t_{i+m} - t_{i+1}} \mathbf{B}_{i+1}^{(m-1)} \quad m = 2, 3. \quad [4.23]$$

En notación matricial, la expresión [4.20] se escribe como:

$$\mathbf{e}_{o(l)}^* = \mathbf{B}_{(l)} \mathbf{b}_{(l)} + \mathbf{e}_{(l)} \quad [4.24]$$

siendo $\mathbf{e}_{o(l)}^*$ el vector de dimensión $(n_l \times 1)$ conteniendo los residuales de los animales de la cabaña l ; $\mathbf{B}_{(l)}$ es una matriz de orden $(n_l \times nx)$ que contiene los vectores columnas $\mathbf{B}_i^{(k)}$, y $\mathbf{b}_{(l)}$ es el vector $(nx \times 1)$ de los parámetros \mathbf{b}_i . Cada fila de $\mathbf{B}_{(l)}$ tiene sus elementos iguales a cero, excepto en las columnas $i - 3$, $i - 2$, $i - 1$ e i donde sus

elementos son los coeficientes básicos $B_{i-3}^{(3)}$, $B_{i-2}^{(3)}$, $B_{i-1}^{(3)}$ y $B_i^{(3)}$, respectivamente. Por lo tanto, cada valor de la covariable DN se transforma en coeficientes *B-spline* en el intervalo (0, 1) para cada animal que pertenezca a un establecimiento con animales nacidos por trasplante embrionario, y posea registro fenotípico en y .

4.3.3.3. Penalizaciones

Dado que los análisis se realizarán independientemente en cada rodeo, abandonaremos el uso del subíndice l para simplificar la presentación. En el caso de residuales distribuidos en forma normal como los presentados en [4.24], los valores de los coeficientes b se estiman de modo que se minimice la suma de los cuadrados de los residuos (S):

$$S = (e_o^* - Bb)' (e_o^* - Bb) \quad [4.25]$$

Por lo tanto, se llega a las ecuaciones:

$$\hat{b} = (B'B)^{-1} B'e_o^* \quad [4.26]$$

y la curva ajustada $\hat{F}_t = B\hat{b}$ depende del tamaño de la base B . Cuanto mayor sea el número de nodos en la base, menos suave es la curva. Si el número de nodos coincide con el número de datos, se obtiene entonces la curva que interpola los datos. Para evitar este problema, O'Sullivan (1986, 1988) introdujo una penalización en la segunda derivada de la curva, de modo que la función objetivo es igual a:

$$S = (e_o^* - Bb)' (e_o^* - Bb) + \lambda_s \int_{t_1}^{t_{nx}} (B''b)^2 dt \quad [4.27]$$

Si bien no hay ninguna particularidad en la segunda derivada y se pueden utilizar derivadas de cualquier orden, este tipo de penalización es bastante común (es la que se utiliza en los *splines* de suavizado) desde la investigación de Reinsch (1967). En el contexto de *splines* de suavizado, la primera derivada conduce a ecuaciones simples y a un ajuste lineal por tramos mientras que, las derivadas de segundo orden conducen a sistemas de ecuaciones más complejos pero con un ajuste más suave. La novedad que introducen las *P-splines* es que la penalización es discreta y que se penalizan los coeficientes directamente, en vez de penalizar la curva, lo que reduce de manera notable la dimensión del problema.

Eilers y Marx (1996) utilizan una penalización basada en la diferencia de orden d entre los coeficientes adyacentes de las bases de *B-splines*. Este tipo de penalización es más flexible ya que es independiente del grado del polinomio utilizado para construir las *B-splines*. Además, es una buena aproximación discreta a la integral de la d -ésima derivada al cuadrado:

$$S = (e_o^* - Bb)' (e_o^* - Bb) + \lambda_s b'D'Db \quad [4.28]$$

En [4.28], λ_s es un escalar positivo que modifica la función de verosimilitud introduciendo una penalización basada en diferencias de orden d entre coeficientes adyacentes. Básicamente, es el escalar que fuerza a los coeficientes a que sigan un patrón suave. A su vez, dicha penalización tiene la particularidad de influenciar solamente a la diagonal principal y a los k elementos próximos a dicha diagonal (en ambos lados de la diagonal principal) del sistema de ecuaciones, quedando así una estructura diagonal en bandas. Bajo este enfoque, el estimador penalizado de \mathbf{b} es la solución al siguiente sistema de ecuaciones:

$$\hat{\mathbf{b}} = (\mathbf{B}'\mathbf{B} + \lambda_s \mathbf{b}'\mathbf{D}'\mathbf{D}\mathbf{b})^{-1} \mathbf{B}'\mathbf{e}_o^* \quad [4.29]$$

Obsérvese que cuando $\lambda_s = 0$, se está en presencia de las ecuaciones de regresión lineal con *B-splines* básicas mientras que, cuando $k = 0$, se produce la situación particular de regresión “ridge”:

$$\mathbf{S} = (\mathbf{e}_o^* - \mathbf{B}\mathbf{b})'(\mathbf{e}_o^* - \mathbf{B}\mathbf{b}) + \lambda_s \mathbf{b}'\mathbf{b} \quad [4.30]$$

En general se utiliza $d = 2$, aunque se pueden emplear órdenes superiores o inferiores, dependiendo de la variabilidad de la curva y la cantidad de “ruido” presente en los datos. Si se considera una diferencia de 1 entre b_i adyacentes ($d = 1$), la suma de cuadrados es:

$$(b_1 - b_2)^2 + \dots + (b_{nx-2} - b_{nx-1})^2 = \sum_{i=1}^{nx} (b_i - b_{i+1})^2 = \mathbf{b}'\mathbf{D}'\mathbf{D}\mathbf{b} \quad [4.31]$$

donde \mathbf{D} es una matriz de orden $(nx-1 \times nx)$. En consecuencia, $\mathbf{D}'\mathbf{D}$ es de orden $(nx \times nx)$. Si, en cambio, se considera $d = 2$, la suma de cuadrados de segundas diferencias es:

$$(b_1 - 2b_2 + b_3)^2 + \dots + (b_{nx-2} - 2b_{nx-1} + b_{nx})^2 = \sum_{i=1}^{nx} (b_i - 2b_{i+1} + b_{i+2})^2 = \mathbf{b}'\mathbf{D}'\mathbf{D}\mathbf{b} \quad [4.32]$$

En este caso \mathbf{D} es de orden $(nx-2 \times nx)$. Nótese que, tanto en [4.31] como en [4.32], $\mathbf{D}'\mathbf{D}$ es de orden $(nx \times nx)$ y es singular. Por ejemplo para $n = 4$, las matrices de primer orden ($d = 1$) \mathbf{D} y $\mathbf{D}'\mathbf{D}$ son respectivamente iguales a:

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}; \quad \mathbf{D}'\mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad [4.33]$$

mientras que, las matrices de segundo ($d = 2$) orden son:

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}; \quad \mathbf{D}'\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ -2 & 5 & -4 & 1 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix} \quad [4.34]$$

Ruppert et al. (2003) y Wand (2003) discutieron la conexión entre *P-splines* y modelos mixtos (Henderson, 1984). En este trabajo se utilizó el enfoque de Cantet et al. (2005), en donde el coeficiente de suavizamiento λ_s resulta del cociente entre las varianzas del error y de los coeficientes *B-splines* presentes en \mathbf{b} . Asimismo, $\mathbf{D'D}$ representa la inversa de la matriz de (co)varianzas de los coeficientes *B-splines*, en lugar de una matriz de penalizaciones. Para mayor detalles sobre el enfoque, véase Cantet et al. (2005).

4.3.3.4. Suavizado de los residuales en rodeos Brangus argentinos.

Los coeficientes de las bases *B-splines* fueron calculados mediante el algoritmo recursivo de De Boor (1993): ver sección 4.3.3.2 del presente capítulo. El número de nodos ajustado fue 43 en todos los establecimientos. Si bien se analizó la posibilidad de cambiar el número de nodos, de modo que queden distanciados cada 3, 6 y 12 meses, su implementación fue imposible dado la falta de años de observación, y, en el caso de períodos cortos, por la presencia de nodos sin registros fenotípicos. La bibliografía concuerda en el uso de un número fijo de 40 nodos (Rupert y Carroll, 2000; Cantet et al., 2005; Durbán, 2009). En el caso de usar menos nodos, existe una tendencia a subajustar (*underfitting*). Por otra parte, el aumento del número de nodos en más de 40 ha demostrado ser marginalmente eficaz en cuanto a la eficiencia de ajuste y la curva tiende suavizarse excesivamente (Eilers and Marx, 1996). En el Cuadro 4.3 se muestra el número de días julianos que abarca un nodo por establecimiento.

Cuadro 4.3: Extensión de los nodos por establecimiento.

Establecimiento	Nº de Días Julianos /Nodo
I	124
II	96
III	79
IV	66
V	246
VI	62
VII	51
VIII	65
IX	114
X	71
XI	73
XII	50

El Cuadro 4.3 refleja la variabilidad en el largo de las series analizadas. Así los nodos están distanciados en menos de 2 meses para los Establecimientos VII y XII mientras que, en el V se encuentran distanciados en más de 8 meses.

Para determinar si los datos tienen alguna característica de autocorrelación, tendencia o estacionalidad se empleó el análisis de Box y Jenkins (1976). A través de una metodología iterativa, este enfoque tiene por objetivo ajustar a la serie de tiempo el modelo autoregresivo (AR), el de medias móviles (MA), de procesos mixtos (ARMA) o procesos integrados (ARIMA), que mejor se adecue a los datos, para posteriormente generar pronósticos. El procedimiento consistió de las siguientes etapas:

1. *Identificación y Selección del modelo.* En primer lugar se evaluó la hipótesis de que las series de tiempo resultantes fuesen “estacionarias” (Hamilton, 1994). Es decir, se examinó si existen evidencias de que los datos poseen media y varianza constantes en el tiempo, y si el valor de la covarianza entre dos períodos depende solamente de la distancia o rezago entre estos dos períodos, y no es función del momento en el cual se ha calculado la covarianza (Hamilton, 1994; Enders, 1995, páginas 68 y 69). Con el objetivo de detectar los componentes esenciales de la serie, es decir, identificar tendencias, variaciones estacionales y variaciones irregulares, también llamada componente aleatoria, se graficaron las series de tiempo por establecimiento. Adicionalmente, se complementaron los análisis gráficos de las 12 series mediante la prueba de Dickey Fuller (DF, Dickey y Fuller, 1979) o el Test de Dickey Fuller Aumentado (ADF, Dickey y

Fuller, 1981; Hamilton, 1994), en función de que el modelo sea AR(1) o que esté correlacionado a un nivel mayor de rezagos. Ambas pruebas detectan la presencia de raíces unitarias, y determinan el grado de una posible diferenciación para lograr la estacionalidad de la serie. La prueba comienza por suponer que la serie b_t (constituida por el BLUP de los nodos en este caso) sigue un proceso AR(1) de la forma:

$$b_t = \theta b_{t-1} + e_t ; \quad |\theta| \leq 1 \quad [4.35]$$

y se realiza la prueba de hipótesis $H_0: \theta = 1$ (de aquí la expresión raíz unitaria). Si el coeficiente de b_{t-1} es igual a 1, se está en una situación de no estacionalidad. Tal como lo indican Dickey y Fuller (1979), se consideraron tres ecuaciones de regresión diferentes para probar la presencia de raíz unitaria:

$$\Delta b_t = (\theta - 1) b_{t-1} + e_t = \delta b_{t-1} + e_t \quad [4.36]$$

$$\Delta b_t = \gamma_1 + \delta b_{t-1} + e_t \quad [4.37]$$

$$\Delta b_t = \gamma_1 + \gamma_2 t + \delta b_{t-1} + e_t \quad [4.38]$$

donde Δ es el operador de primera diferencia estacionaria, y t representa la variable tiempo o tendencia. El primero es un modelo puro de caminata aleatoria, el segundo agrega un intercepto o término de deriva, y el tercero incluye ambos, es decir un intercepto y una línea de tendencia. En cada caso, la hipótesis nula de interés es que $\delta = 1$, es decir, que hay una raíz unitaria. Si el término de error está autocorrelacionado, se modifica [4.38] de modo de contemplar términos autorregresivos de mayor orden. Esta prueba se conoce como el test de Dickey-Fuller Aumentado:

$$\Delta b_t = \gamma_1 + \gamma_2 t + \delta b_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta b_{t-i} + e_t \quad [4.39]$$

En [4.39] se utilizan términos en diferencia rezagados, los que serán incluidos hasta que el término de error no contenga autocorrelación. Las pruebas de hipótesis son las mismas que en la prueba de DF. Luego, se analizaron las funciones de autocorrelación simple (ACF) y las de autocorrelación parcial (PACF). Se seleccionaron los modelos candidatos a evaluarse observando las gráficas ACF y PACF de la serie. Las autocorrelaciones de una serie totalmente aleatoria (“ruido blanco”) oscilan alrededor del cero. El número máximo de valores anteriores (“*lag length*”) considerado fue 10, dado que el “*lag length*” es de $n/4$ cuando la serie cuenta con menos de 240 datos (SAS Institute Inc. 2008, página 229).

2. *Estimación de los parámetros.* Una vez identificados los modelos candidatos, se estimaron los parámetros de dispersión mediante el método REML (*Restricted Maximum Likelihood*, Patterson y Thompson, 1971) dentro del PROC Mixed de SAS, y se evaluó la significancia de los parámetros. Luego, se seleccionó el modelo con los valores más bajos del criterio AIC (Criterio de Información de Akaike). Finalmente, se ajustó a los residuales de los pesos al destete, el modelo con *P-splines* y una matriz de covarianzas siguiendo un proceso AR(1).

3. *Verificación de la adecuación del modelo.* Particularmente, se constató que los residuales del modelo fuesen independientes entre sí. Para ello, se realizó el test de “ruido blanco” de Ljung y Box (1978), empleando las estimaciones de los coeficientes *B-splines*. Se espera no rechazar la hipótesis nula que soporta la presencia de “ruido blanco”. De lo contrario, se concluirá que existen errores sistemáticos que deben ser tenidos en cuenta por el modelo, volviéndose a la etapa 2 anterior.

Los parámetros obtenidos serán utilizados en el cálculo de las predicciones BLUP con un modelo animal modificado acorde a la nueva estructura del error, hecho que se describirá en la sección 4.5.

4.4. RESULTADOS

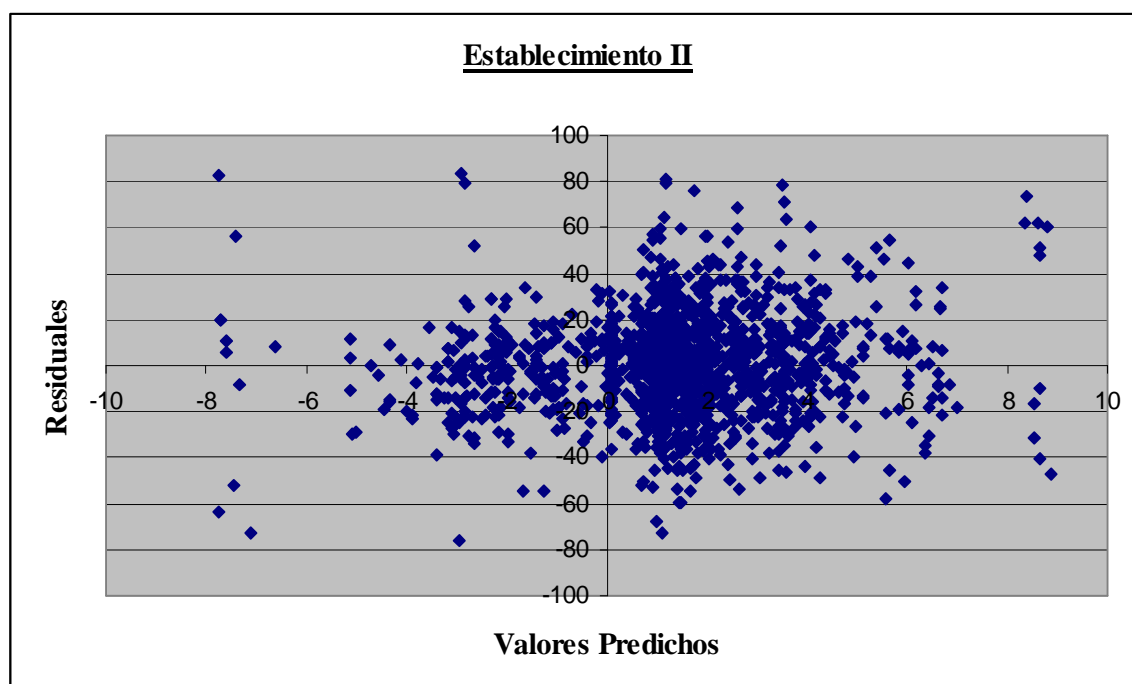
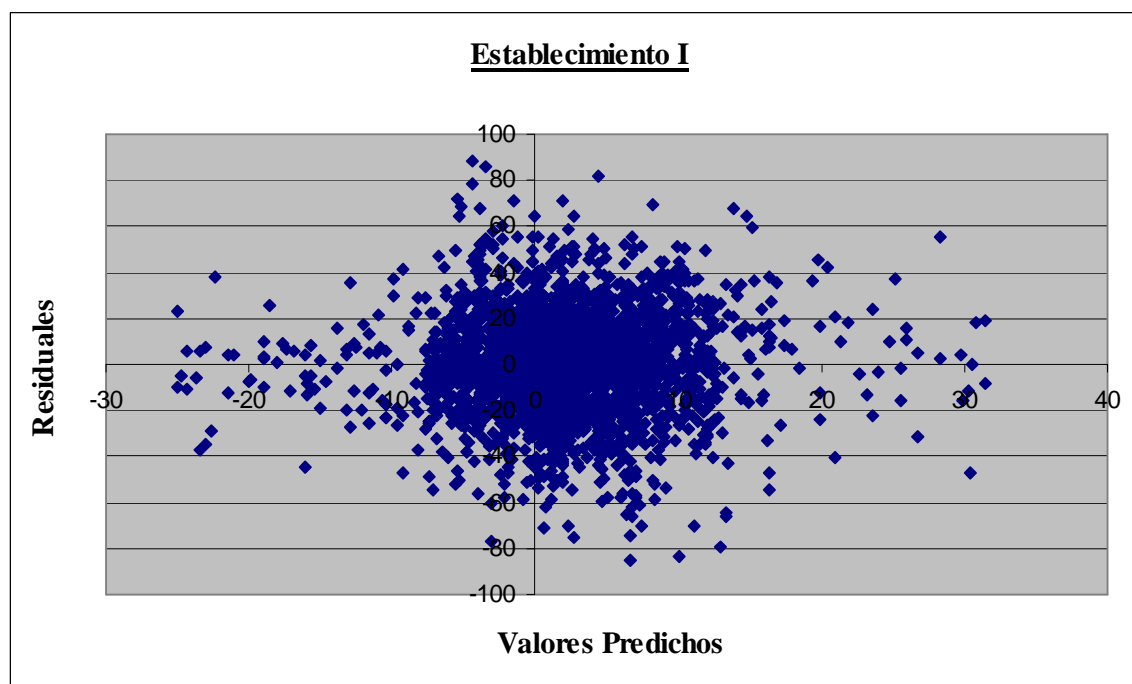
En los 12 casos se rechazó la hipótesis nula de los test de DF y ADF. Con este resultado se concluyó que no existen “raíces unitarias”, ni es necesario integrar las series. En el Cuadro 4.4 se muestran las estimaciones de los parámetros de dispersión de los modelos propuestos. En 10 de los 12 establecimientos se pudo ajustar la estructura AR(1), representada por los parámetros: σ_b^2 y ρ . En los establecimientos V y VII se ajustó un modelo con estructura esférica ($I \sigma_b^2$). Posiblemente el menor número de datos del rodeo VII hayan sido la razón de la falta de ajuste. En el establecimiento V, el valor estimado del parámetro ρ fue muy cercano a 0, considerándose así una estructura alternativa $I \sigma_b^2$.

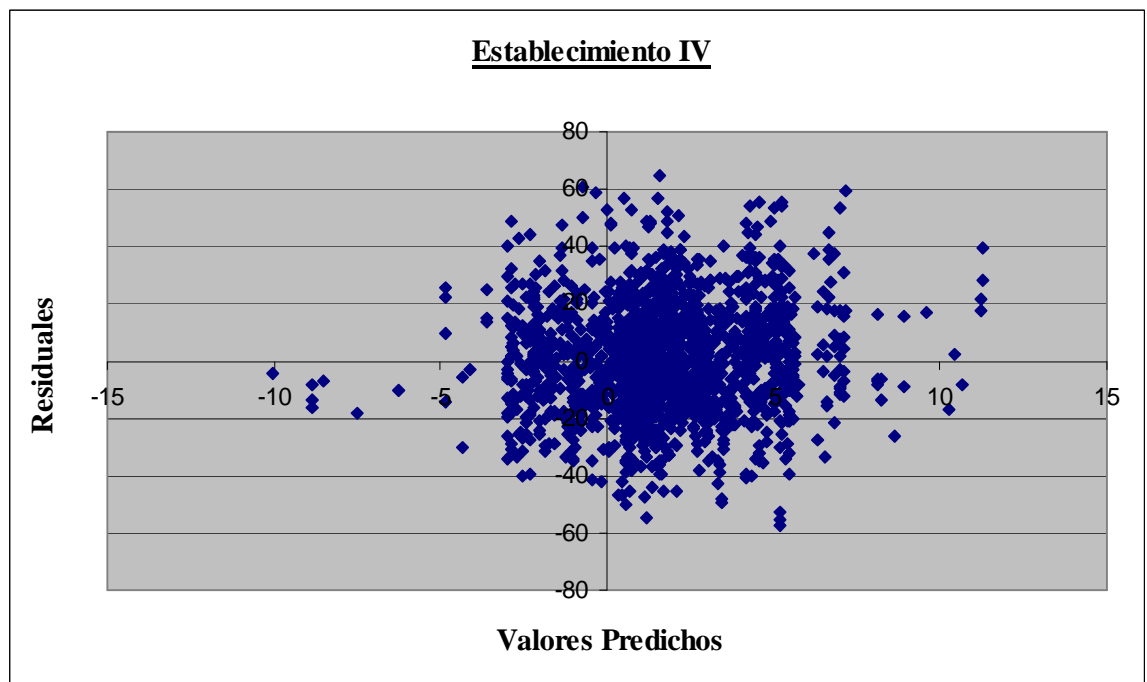
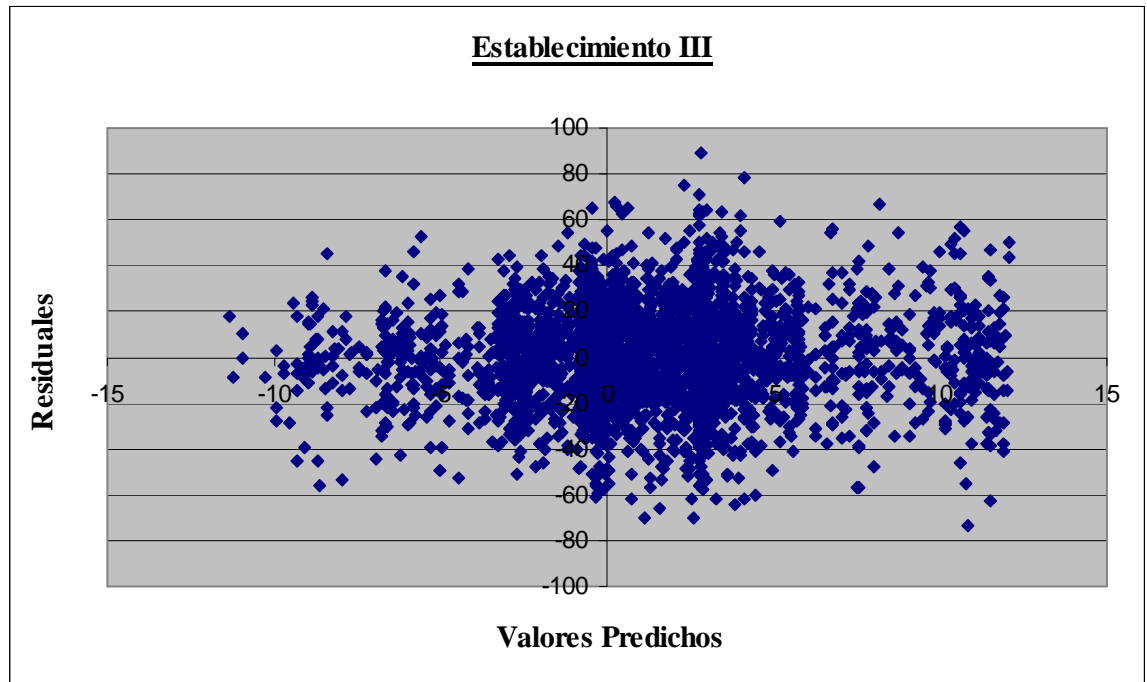
Cuadro 4.4: Estimaciones de los componentes de (co)varianza por cabaña.

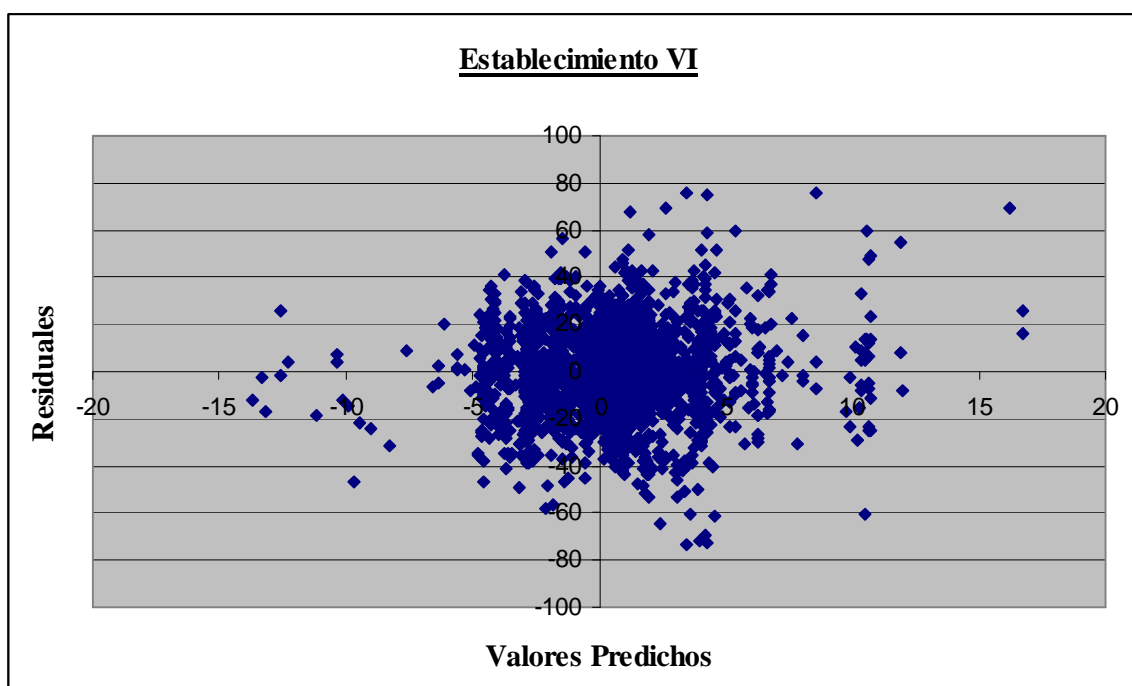
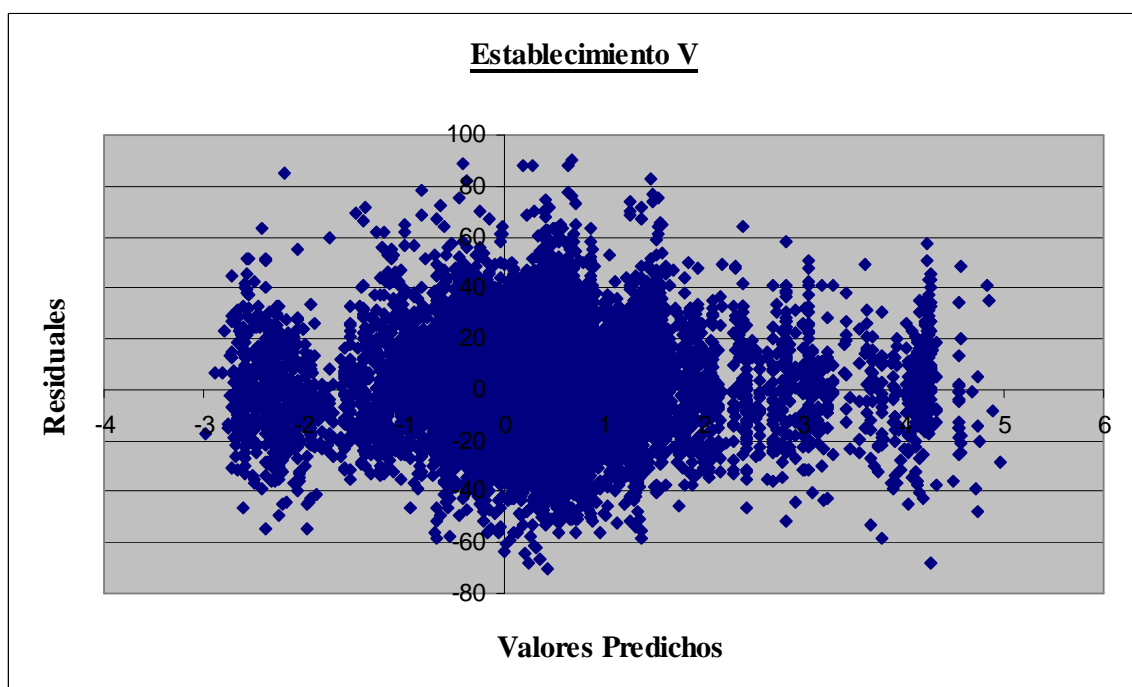
Establecimiento	σ_b^2	ρ	σ_e^2
I	708.15	-0.66	453.69
II	52.75	-0.39	480.96
III	204.16	-0.75	460.89
IV	115.49	-0.73	358.41
V	7.06	—	331.47
VI	117.53	0.15	354.70
VII	12.90	—	512.25
VIII	60.17	-0.52	429.60
IX	171.18	-0.61	278.29
X	85.73	-0.24	452.68
XI	79.90	-0.56	242.95
XII	129.13	-0.35	382.16

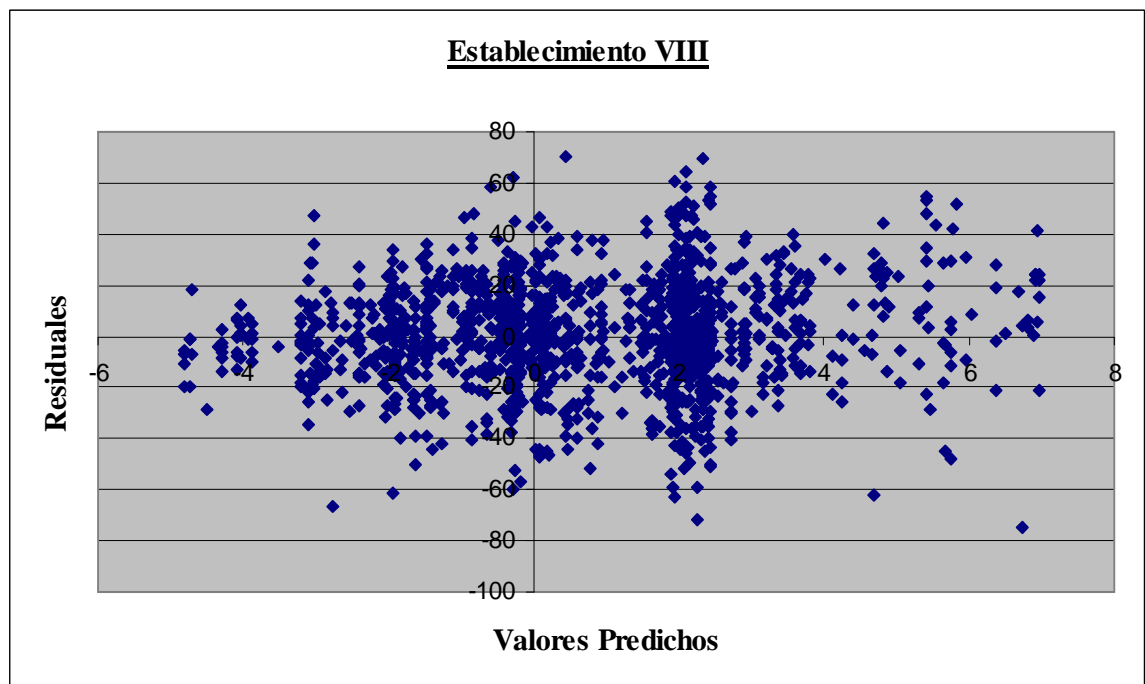
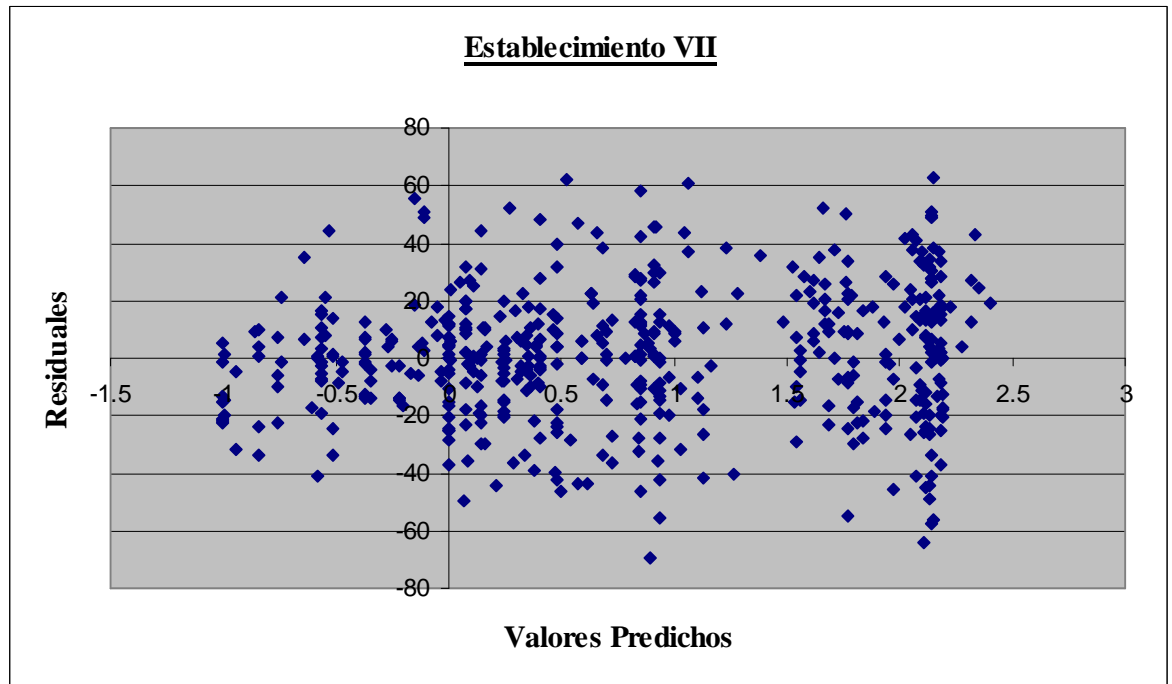
De acuerdo con los resultados presentados en el Cuadro 4.4, en 10 de los 12 establecimientos, el modelo AR(1) tuvo un buen comportamiento. Salvo en el establecimiento VI, el parámetro ρ fue negativo. Al verificar la adecuación del modelo, se graficaron los residuales vs los valores predichos de los 12 establecimientos analizados. Se buscó que los residuos indiquen un proceso aleatorio. En la Figura 4.2 se muestran los gráficos resultantes.

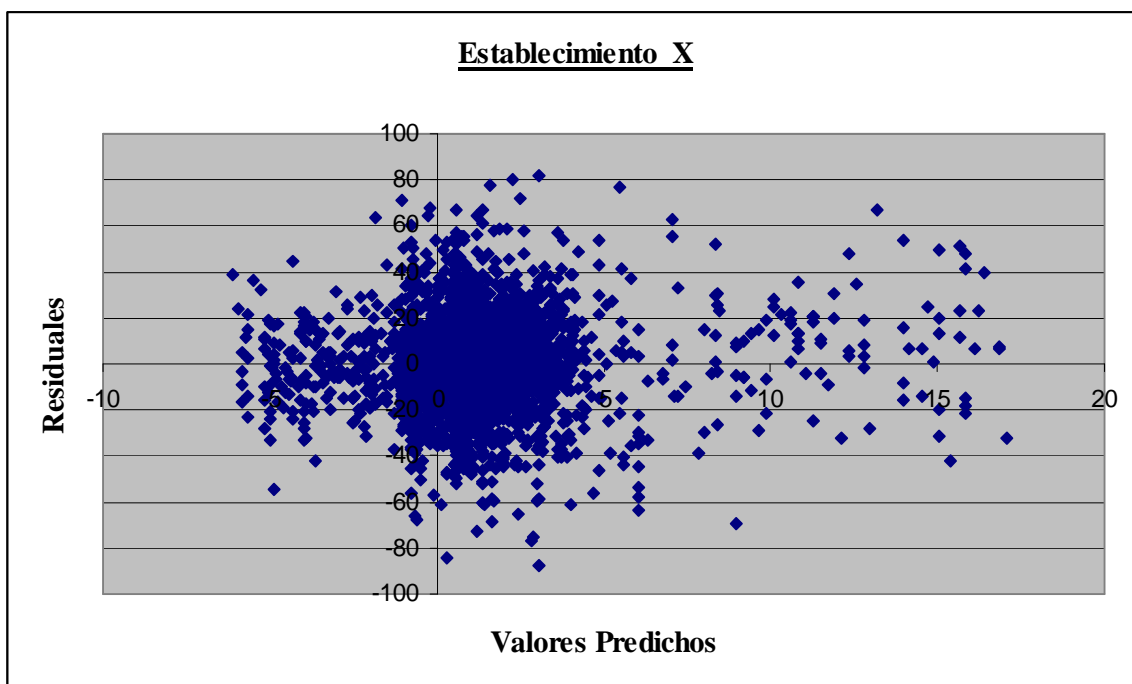
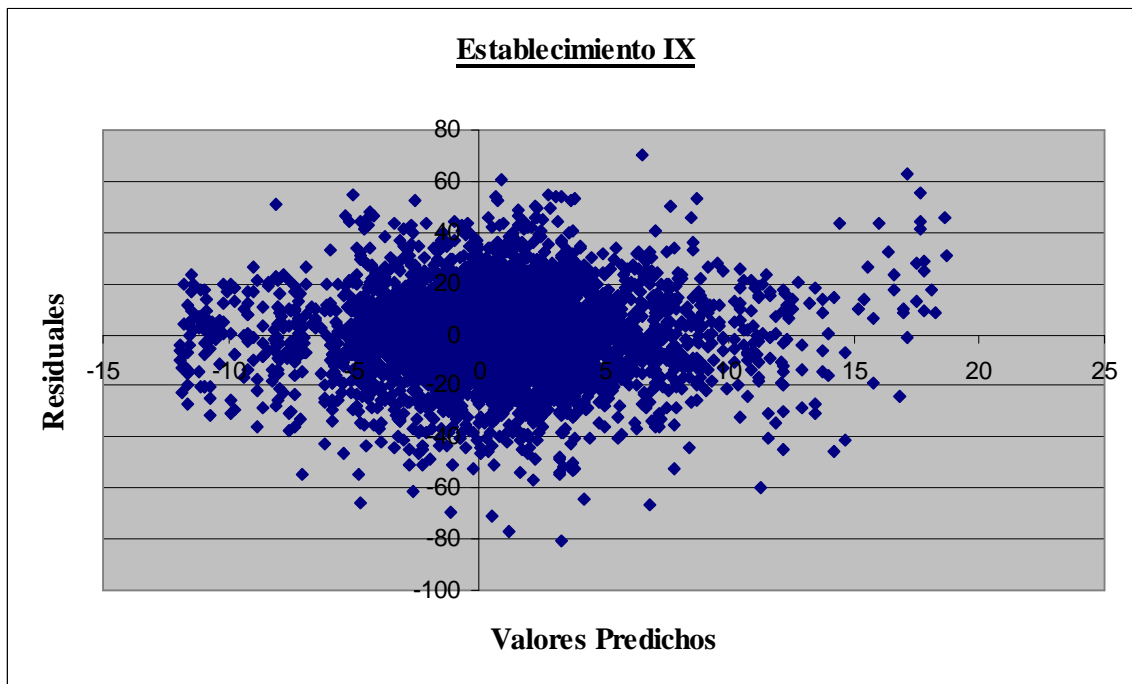
Figura 4.2: Gráficos de Residuales vs Valores Predichos de los 12 establecimientos analizados.

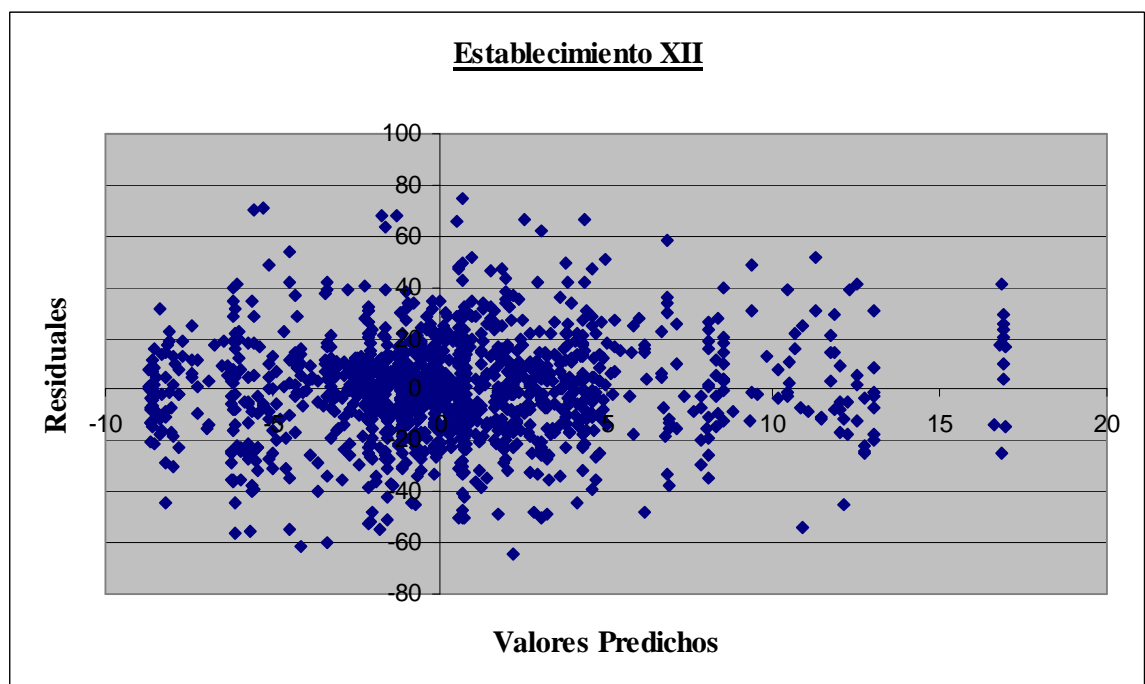
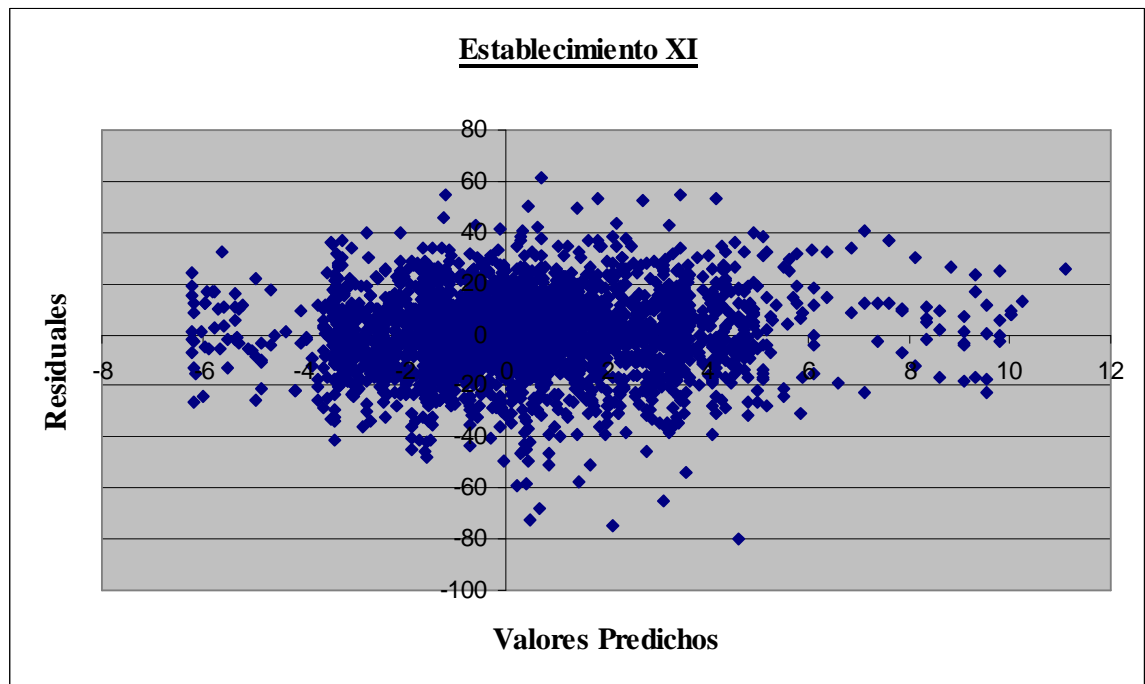












En referencia a los gráficos anteriores, se puede observar que los datos se encuentran distribuidos aleatoriamente, y que no hay evidencias de algún patrón definido. Estos análisis, a su vez, fueron complementados con las pruebas de ruido blanco de Ljung y Box (1978).

4.5. IMPLEMENTACIÓN EN LA EVALUACIÓN GENÉTICA

Al llevar a cabo la evaluación genética de animales nacidos por transplante embrionario se pueden presentar dos situaciones: 1) que la información de las madres biológicas y receptoras esté completa; o 2) que las madres receptoras no estén identificadas. En el primer caso, el modelo a ajustar corresponde con el “modelo de tres padres” presentado por Van Vleck (1990b) y Schaeffer y Kennedy (1989). En dicho modelo, los efectos genéticos corresponden a la madre biológica mientras que, los efectos maternos, los efectos ambientales maternos permanentes y el efecto fijo EM son provistos por la vaca receptora. Si bien el modelo contempla la variabilidad existente entre razas en lo que respecta a los efectos maternos y al ambiente materno permanente, en la presente investigación dicha variabilidad no fue tomada en cuenta porque se asumió que todas las vacas receptoras pertenecen a la misma raza que los terneros a evaluar, sobre la base de la información suministrada por los criadores. En el segundo caso, debido a la imposibilidad de contar con la información completa de la vaca receptora, el modelo a ajustar debe corregir el sesgo debido al error de medición clásico generado por la EM incierta; incrementar la varianza residual en $\beta_2^2 \sigma_u^2 \lambda$ unidades, y agregar un término adicional correspondiente a los coeficientes *B-splines*. De este modo, la ecuación del modelo para los animales nacidos por transplante embrionario cuya madre receptora es desconocida resulta igual a:

$$y_{ij} = \mathbf{x}_{1i} \boldsymbol{\beta}_1 + w_i \beta_2 + a_{oi} + a_{mj} + e_{mj} + \mathbf{B}_{i(l)} \mathbf{b}_{i(l)} + e_{oi} \quad [4.40]$$

El subíndice $l = 1, \dots, n_e$ se utiliza para identificar los distintos rodeos. Por ejemplo, en los animales Brangus aquí analizados $n_e = 12$. A su vez, si se agrupan las observaciones correspondientes a todos los animales pertenecientes al rodeo l -ésimo, la expresión matricial de [4.40] es la siguiente:

$$\mathbf{y}_{(l)} = \mathbf{X}_{1(l)} \boldsymbol{\beta}_1 + \mathbf{w}_{(l)} \beta_2 + \mathbf{Z}_{o(l)} \mathbf{a}_{o(l)} + \mathbf{Z}_{m(l)} \mathbf{a}_{m(l)} + \mathbf{Z}_{p(l)} \mathbf{e}_{m(l)} + \mathbf{B}_{(l)} \mathbf{b}_{(l)} + \mathbf{e}_{o(l)} \quad [4.41]$$

Los términos de las ecuaciones [4.40] y [4.41] son los mismos que los presentados en las secciones 4.3.3.1 y 4.2.1.3.

En los establecimientos que no poseen transplantes embrionarios entre sus registros, los modelos a ajustar son: a) el modelo animal con efectos maternos (Quaas y Pollak, 1980, ver capítulo 2):

$$\mathbf{y}_{(l)} = \mathbf{X}_{1(l)} \boldsymbol{\beta}_1 + \mathbf{Z}_{o(l)} \mathbf{a}_{o(l)} + \mathbf{Z}_{m(l)} \mathbf{a}_{m(l)} + \mathbf{Z}_{p(l)} \mathbf{e}_{m(l)} + \mathbf{e}_{o(l)} \quad [4.42]$$

ó b) el modelo animal con error de medición clásico en el efecto fijo EM (ver sección 4.2.1.3):

$$\mathbf{y}_{(l)} = \mathbf{X}_{1(l)} \boldsymbol{\beta}_1 + \mathbf{w}_{(l)} \beta_2 + \mathbf{Z}_{o(l)} \mathbf{a}_{o(l)} + \mathbf{Z}_{m(l)} \mathbf{a}_{m(l)} + \mathbf{Z}_{p(l)} \mathbf{e}_{m(l)} + \mathbf{e}_{o(l)} \quad [4.43]$$

en función que las madres de los terneros estén o no identificadas.

La matriz \mathbf{B} es diagonal en bloques, y presenta la siguiente forma:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{(1)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{(2)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_{(3)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_{(n_e)} \end{bmatrix} \quad [4.44]$$

Cada bloque corresponde a una submatriz $\mathbf{B}_{(l)}$ de orden $(n_l \times nx)$ que contiene los vectores columnas $\mathbf{B}_i^{(k)}$. A su vez, cada fila de $\mathbf{B}_{(l)}$ tiene sus elementos iguales a cero, excepto en las columnas $i-3$, $i-2$, $i-1$ e i donde sus elementos son los coeficientes básicos $\mathbf{B}_{i-3}^{(3)}$, $\mathbf{B}_{i-2}^{(3)}$, $\mathbf{B}_{i-1}^{(3)}$ y $\mathbf{B}_i^{(3)}$, respectivamente. El vector \mathbf{b} es el siguiente:

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_{(1)} \\ \mathbf{b}_{(2)} \\ \mathbf{b}_{(3)} \\ \vdots \\ \mathbf{b}_{(n_e)} \end{bmatrix} \quad [4.45]$$

donde $\mathbf{b}_{(l)}$ es un subvector $(nx \times 1)$ de las soluciones de las *B-splines* (\mathbf{b}) para el rodeo l . Se asume que todos los vectores aleatorios en [4.41]-[4.43] tienen esperanza cero, y estructura de (co)varianzas igual a:

$$\mathbb{E} \begin{bmatrix} \mathbf{a}_o \\ \mathbf{a}_m \\ \mathbf{e}_m \\ \mathbf{b} \\ \mathbf{e}_o \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; \quad \text{Var} \begin{bmatrix} \mathbf{a}_o \\ \mathbf{a}_m \\ \mathbf{e}_m \\ \mathbf{b} \\ \mathbf{e}_o \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_{Ao}^2 & \mathbf{A}\sigma_{AoAm} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{A}\sigma_{AoAm} & \mathbf{A}\sigma_{Am}^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_d \sigma_{Em}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix} \quad [4.46]$$

donde la matriz \mathbf{S} es igual a:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{(1)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{(2)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{(3)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_{(n_e)} \end{bmatrix} \quad [4.47]$$

Cada bloque $\mathbf{S}_{(l)}$ representa la siguiente estructura:

$$\mathbf{S}_{(l)} = \mathbf{M}_{(l)} \sigma_{b(l)}^2 \quad [4.48]$$

El escalar $\sigma_{b(l)}^2$ es la varianza entre los coeficientes *B-splines* correspondiente al l -ésimo establecimiento, mientras que la matriz $\mathbf{M}_{(l)}$ representa la estructura autoregresiva de orden 1 que se muestra a continuación:

$$\mathbf{M}_{(l)} = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots \\ \rho & 1 & \rho & \rho^2 & \cdots \\ \rho^2 & \rho & 1 & \rho & \cdots \\ \rho^3 & \rho^2 & \rho & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad [4.49]$$

El parámetro ρ tiene la restricción: $|\rho| < 1$. Obsérvese que [4.49] tiene un rol importante dentro del modelo [4.41] dado que refleja la caída en la correlación entre pares de nodos que se encuentran cada vez más alejados. Debido a la falta de información en los datos, para algunos establecimientos la matriz \mathbf{M} debe reemplazarse por una matriz identidad. En estos casos tenemos:

$$\mathbf{S}_{(l)} = \mathbf{I}_{(n_l)} \sigma_{b(l)}^2 \quad [4.50]$$

La estructura de (co)varianzas del error también es diagonal en bloques:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{(1)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{(2)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_{(3)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_{(n_e)} \end{bmatrix} \quad [4.51]$$

En los rodeos con animales nacidos por transplantes embrionarios o con madres desconocidas, la submatriz $\mathbf{R}_{(l)}$ es igual a:

$$\mathbf{R}_{(l)} = \begin{bmatrix} \mathbf{I}_{n_{C(l)}} \sigma_{e(l)}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_{D(l)}} [\beta_2^2 \sigma_u^2 \lambda + \sigma_{e(l)}^2] \end{bmatrix} \quad [4.52]$$

mientras que, en los rodeos con toda la información de las vacas biológicas y/o receptoras completa, la submatriz $\mathbf{R}_{(l)}$ es igual a:

$$\mathbf{R}_{(l)} = \mathbf{I}_n \sigma_{\text{Eo}}^2 \quad [4.53]$$

En consecuencia, al implementarse la evaluación genética, las ecuaciones de modelo mixto a resolver son las siguientes:

$$\begin{bmatrix} \mathbf{X}_1' \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{X}_1' \mathbf{R}^{-1} \mathbf{w} \lambda^{-0.5} & \mathbf{X}_1' \mathbf{R}^{-1} \mathbf{Z}_o & \mathbf{X}_1' \mathbf{R}^{-1} \mathbf{Z}_m & \mathbf{X}_1' \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{X}_1' \mathbf{R}^{-1} \mathbf{B} \\ \lambda^{-0.5} \mathbf{w}' \mathbf{R}^{-1} \mathbf{X}_1 & \lambda^{-1} \mathbf{w}' \mathbf{R}^{-1} \mathbf{w} & \lambda^{-0.5} \mathbf{w}' \mathbf{R}^{-1} \mathbf{Z}_o & \lambda^{-0.5} \mathbf{w}' \mathbf{R}^{-1} \mathbf{Z}_m & \lambda^{-0.5} \mathbf{w}' \mathbf{R}^{-1} \mathbf{Z}_p & \lambda^{-0.5} \mathbf{w}' \mathbf{R}^{-1} \mathbf{B} \\ \mathbf{Z}_o' \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{Z}_o' \mathbf{R}^{-1} \mathbf{w} \lambda^{-0.5} & \mathbf{Z}_o' \mathbf{R}^{-1} \mathbf{Z}_o + \mathbf{A}^{-1} \mathbf{g}^{11} & \mathbf{Z}_o' \mathbf{R}^{-1} \mathbf{Z}_m + \mathbf{A}^{-1} \mathbf{g}^{12} & \mathbf{Z}_o' \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{Z}_o' \mathbf{R}^{-1} \mathbf{B} \\ \mathbf{Z}_m' \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{Z}_m' \mathbf{R}^{-1} \mathbf{w} \lambda^{-0.5} & \mathbf{Z}_m' \mathbf{R}^{-1} \mathbf{Z}_o + \mathbf{A}^{-1} \mathbf{g}^{21} & \mathbf{Z}_m' \mathbf{R}^{-1} \mathbf{Z}_m + \mathbf{A}^{-1} \mathbf{g}^{22} & \mathbf{Z}_m' \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{Z}_m' \mathbf{R}^{-1} \mathbf{B} \\ \mathbf{Z}_p' \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{Z}_p' \mathbf{R}^{-1} \mathbf{w} \lambda^{-0.5} & \mathbf{Z}_p' \mathbf{R}^{-1} \mathbf{Z}_o & \mathbf{Z}_p' \mathbf{R}^{-1} \mathbf{Z}_m & \mathbf{Z}_p' \mathbf{R}^{-1} \mathbf{Z}_p + \mathbf{I}_d \frac{1}{\sigma_{\text{Em}}^2} & \mathbf{Z}_p' \mathbf{R}^{-1} \mathbf{B} \\ \mathbf{B}' \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{B}' \mathbf{R}^{-1} \mathbf{w} \lambda^{-0.5} & \mathbf{B}' \mathbf{R}^{-1} \mathbf{Z}_o & \mathbf{B}' \mathbf{R}^{-1} \mathbf{Z}_m & \mathbf{B}' \mathbf{R}^{-1} \mathbf{Z}_p & \mathbf{B}' \mathbf{R}^{-1} \mathbf{B} + \mathbf{S}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\alpha}_o \\ \hat{\alpha}_m \\ \hat{\epsilon}_m \\ \hat{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1' \mathbf{R}^{-1} \mathbf{y} \\ \lambda^{-0.5} \mathbf{w}' \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_o' \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_m' \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}_p' \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{B}' \mathbf{R}^{-1} \mathbf{y} \end{bmatrix} \quad [4.54]$$

donde $\begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix} = \begin{bmatrix} \sigma_{Ao}^2 & \sigma_{AoAm} \\ \sigma_{AoAm} & \sigma_{Am}^2 \end{bmatrix}^{-1}$. La inversión de S es computacionalmente sencilla al emplear el algoritmo presentado por Wade y Quaas (1993, p.3030).

4.6. DISCUSIÓN

En este capítulo se propusieron dos soluciones al problema de la evaluación genética cuando en los datos de los animales nacidos por trasplante embrionario no se identifica a las madres receptoras. En primer lugar se abordó el problema de la edad de madre incierta mediante la inclusión de errores de medición clásicos en el efecto fijo EM. Además, se detectó la presencia de varianza residual mediante técnicas de series de tiempo equiespaciadas, a través del ajuste de *splines* penalizadas. La primer metodología, que es ampliamente difundida en los estudios de epidemiología, es el primer antecedente en el área de mejoramiento animal para atenuar el sesgo por error de medición clásico en el efecto de EM, cuando el nivel del efecto es desconocido. Schaeffer y Kennedy (1989) y Van Vleck (1990b) trataron la evaluación genética de animales nacidos por trasplante embrionario, y destacaron la importancia de identificar la EM de la hembra receptora, sin embargo, no realizaron ninguna sugerencia sobre cómo tratar EM desconocidas (Van Vleck, 1990b, página 4027). Nótese que este problema subsiste aún cuando se asignen madres “fantasmas” (Van Vleck, 1990a), o cuando se reemplacen los valores de cría maternos de las hembras desconocidas que poseen un hijo con registro fenotípico, con una función lineal del BLUP de su progenie (multiplicada por 2/3) y el padre de su cría (multiplicada por -1/3) (ver capítulo 3). Ambas soluciones permiten especificar correctamente la estructura de (co)varianzas aditivas pero no corrigen el sesgo que se introduce por desconocer la EM. La raza de las receptoras es otra información relevante a considerar pero, en la presente tesis dicho efecto no fue tenido en cuenta porque en las razas analizadas, las vacas receptoras pertenecen a la misma raza que las vacas dadoras.

Por sus características biológicas, los PD de los animales fueron observados en forma secuencial pero no a intervalos regulares de tiempo. Esto impide realizar los análisis clásicos de series de tiempo sobre los residuales, para lo cual se requirió ajustar una función mediante el método de *P-splines* de modo de mapear los residuales a una serie temporal a intervalos de medición regulares. Sin embargo, existieron otras razones que justificaron el uso de esta metodología. Por un lado, las *splines* utilizadas son de bajo rango, es decir, el tamaño de la base fue mucho menor que la dimensión de los datos. Esto contrasta con lo que ocurre, por ejemplo, con las *splines* de suavizado donde hay tantos nodos como datos, lo que provoca que haya que trabajar con matrices de grandes dimensiones (Durbán, 2009, pág. 196). En el caso de las *P-splines*, el número de nodos no supera los 40, con lo cual son computacionalmente eficientes, sobre todo si se trabaja con rodeos con gran cantidad de datos. Otro argumento favorable es la introducción de penalizaciones, hecho que disminuye la importancia de la elección del número y la localización de los nodos (Rice y Wu, 2001), evitando tener que seleccionar el número de nodos en función del establecimiento. Finalmente, la correspondencia entre las *P-splines* y el BLUP permite utilizar la metodología existente en el campo de los modelos mixtos, y el uso de software estadístico como SAS: otro de los puntos claves de la presente investigación. El objetivo fue encontrar, no sólo una estructura de correlación que se adecue a los datos disponibles, sino que también sea factible de incorporar dentro de las ecuaciones de modelo mixto para un modelo animal. Wade y

Quaas (1993) desarrollaron un algoritmo que permite modelar un efecto aleatorio con una estructura de (co)varianzas AR(1) e incorporarlo en el sistema de ecuaciones de modelo mixto. De manera similar a lo que ocurre con la matriz de relaciones de parentesco de Henderson (1976), el algoritmo permite obtener la inversa de la matriz de (co)varianzas con estructura AR(1) sin necesidad de invertir la matriz original, a través del seguimiento de unas reglas sencillas. En un artículo posterior, Wade y colaboradores (1993) mostraron cómo estimar los parámetros de dispersión de los procesos AR(1) mediante REML siguiendo el algoritmo EM.

En el presente capítulo no se estudiaron las causas que generan la correlación entre nodos sucesivos. De hecho, se desconoce si dicha correlación es un artefacto del modelo, o de los múltiples factores que se encuentran afectando a las observaciones más próximas. Hubiese sido óptimo contar con series de tiempo de, por lo menos, 40 años para disponer de nodos espaciados cada 365 días julianos. Series con esas características nos hubieran permitido poner a prueba la hipótesis de que existen covarianzas temporales debido al uso repetido de vacas receptoras.

Capítulo 5

Discusión general

Esta tesis presenta algunas contribuciones teóricas y otras metodológicas, relacionadas con los modelos de evaluación genética con efectos maternos, o ancestrales por vía materna, cuando las madres y/o las abuelas maternas no están identificadas. La temática es especialmente relevante para la evaluación genética de poblaciones con políticas de pedigrí “abierto”. Cabe aclarar que esta situación es la que prima en razas compuestas bovinas como Brangus o Braford, de gran desarrollo en países latinoamericanos. Se debe destacar que el modelo de evaluación genética para caracteres con efectos maternos y animales nacidos de transplante embrionario es debido a Schaeffer y Kennedy (1989) (“modelo de tres padres”). Adicionalmente las contribuciones de Van Vleck (1990a) y Cantet et al. (1992a,b) fueron realizadas sobre la base del MAM clásico, pero no resuelven los problemas de la incorrecta especificación del efecto de edad de la madre y de la varianza residual, de las observaciones provenientes de terneros con madres no identificadas.

El análisis en el capítulo 2 consistió en la búsqueda de un modelo que: 1) explique razonablemente la variabilidad de un carácter con efectos maternos, 2) sea parsimonioso (es decir, que no contenga excesivos parámetros en relación con la información disponible); 3) los componentes de (co)varianza resultantes sean “identificables” (Jiang, 1996); 4) la estimación sea numéricamente factible. El análisis de varios modelos con estructuras de (co)varianza fuertemente parametrizadas mostró que la inclusión de efectos ancestrales por vía materna, sean los valores cría de abuela materna o los efectos ambientales maternos permanentes correlacionados por línea materna, redujeron la magnitud negativa de la correlación entre los efectos aditivos directos y maternos, en una base de datos donde los animales con registros fenotípicos poseían sus abuelas maternas identificadas. La mayor varianza aditiva total estimada y la remarcable disminución de la magnitud de la covarianza σ_{AoAm} , favoreció el ajuste de modelos incluyendo los efectos ambientales maternos permanentes no independientes, dentro de la evaluación genética del peso al destete en bovinos de carne. La ventaja de este modelo es que sólo incorpora un parámetro adicional al MAM tradicional: el coeficiente de auto-regresión dentro de líneas maternas de efectos ambientales permanentes (Quintanilla et al., 1999). La aplicación de estos modelos a bases de datos reales requiere evaluar la estructura de datos en relación con la información de pedigrí. Estructuras de pedigrí incompleto difícilmente justifiquen la sobreparameterización que conllevan los modelos con efectos maternos ancestrales (sean con efectos ambientales maternos permanentes correlacionados o con efectos aditivos de abuela materna), dado que la multicolinealidad de efectos produciría predicciones muy variables, de vacas en distintos años, hecho que suele desacreditar las evaluaciones genéticas entre los criadores. En oposición, pedigríes muy completos (como ocurre en las razas británicas) favorecerían el ajuste de modelos con efectos ancestrales para disminuir el impacto negativo de una magnitud importante en el valor de σ_{AoAm} .

En el capítulo 3, se desarrollaron modelos equivalentes a los presentados por Willham (1972), Cantet et al., (1992b) y Van Vleck (1990a), que permiten especificar correctamente la fracción aditiva de las varianzas y las covarianzas cuando se desconocen las madres y/o las abuelas maternas de los animales con datos, mediante una serie de modificaciones en las líneas de código de los programas de evaluación genética estándar. La gran ventaja del uso de dichos modelos es la marcada disminución en el número de ecuaciones a resolver dado que permite absorber los valores de cría de las madres y abuelas “fantasmas”. Desde el punto de vista computacional, la ganancia en el tiempo de cómputo dependerá del número de animales con registro fenotípico y de la fracción de abuelas maternas desconocidas en la base de datos. Si bien, la velocidad de cálculo y poder de almacenamiento de las computadoras actuales es suficiente como para resolver grandes sistemas de ecuaciones, un evaluador anónimo del trabajo que fuera oportunamente aceptado en *Livestock Science* señaló que “el enfoque propuesto reduce el tiempo de cálculo pero complica un poco el código de los programas de evaluación para estos modelos que no son corrientes”. Y agregó que “el trabajo tiene un evidente interés como formulación de modelos equivalentes; provee un conocimiento más profundo de la relación entre los elementos del modelo y, consecuentemente, una mejor comprensión de los mismos. Por lo tanto, tiene un claro valor pedagógico”.

En el capítulo 4 se presentó un modelo de evaluación genética que atenúa el sesgo en el efecto EM, para los registros de animales con madres desconocidas, y que suaviza los residuales de los datos provenientes de rodeos con datos de transplante embrionario y madres receptoras desconocidas. Si bien la metodología MEM se aplicó sobre la edad de la madre puede ser utilizada para otros efectos fijos, siempre y cuando: 1) los errores de medición sean no diferenciales; 2) el factor de atenuación λ sea estimable. Asimismo, esta metodología puede combinarse con los modelos propuestos por R. J. C. Cantet y R. L. Quaas (presentados en el capítulo 3). Finalmente, y debido a la cantidad de años de registros disponibles, en este capítulo no se pudieron estudiar las causas que generan la correlación entre nodos sucesivos. De hecho, se desconoce si dicha correlación es un artefacto del modelo, o de los múltiples factores que se encuentran afectando a las observaciones más próximas. Resultaría interesante aplicar la metodología a series de tiempo de, por lo menos, 40 años para disponer de nodos espaciados cada 365 días julianos. Series con esas características permitirían poner a prueba la hipótesis inicial de que existen covarianzas temporales debido al uso repetido de vacas receptoras. Cualquiera sea la causa de covariación residual, el análisis de datos reales mostró que el fenómeno parece ser más general de lo que podría esperarse y quizás sea debido a otras razones distintas a la repetición de madres desconocidas entre años. Esta última reflexión sugiere que la investigación metodológica en los modelos de evaluación genética aún tiene problemas para resolver y respuestas para dar.

BIBLIOGRAFÍA CITADA

- Baschnagel, M. B., J. Moll, and N. Künzi. 1999. Comparison of models to estimate maternal effects for weaning weight of Swiss Angus cattle fitting a sire \times herd interaction as an additional random effect. *Livest. Prod. Sci.* 60:203–208.
- Berkson, J. 1950. Are there two regressions? *J. Am. Stat. Assoc.* 45:164–180.
- Bijma, P. 2006. Estimating maternal genetic effects in livestock. *J. Anim. Sci.* 84:800–806.
- Box, G. E. P., and G. C. Tiao. 1973. *Bayesian inference in statistical analysis*. Addison-Wesley Publishing Co, Reading, MA.
- Box, G. E. P., and G. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*, San Francisco, CA: Holden-Day.
- Brown, P. J. 2001. The generalized inverted Wishart distribution. In: El-Sharaawi, A.H. and W. W. Piegorsch Eds. 'Encyclopedia of Environmetrics, volume 2'. Wiley, NY.
- Bulmer, M. G. 1985. *The mathematical theory of quantitative genetics*. 2nd edition. Oxford, London. U.K.
- Buzas, J. S., T. D. Tosteson, and L. A. Stefansky. 2005. Measurement error. *Handbook of epidemiology*. Springer, Berlín:729–765.
- Cantet, R. J. C. 2005. Modelo de error de medición en la evaluación genética con edad de madre incierta. *Rev. Arg. Prod. Anim.* 25 (Supl. 1):GM. Abs.
- Cantet, R. J. C., D. D. Kress, D. C. Anderson, D. E. Doornbos, P. J. Burfening, and R. L. Blackwell. 1988. Direct and maternal variances and covariances and maternal phenotypic effects on preweaning growth of beef cattle. *J. Anim. Sci.* 66:648–660.
- Cantet, R. J. C., R. L. Fernando, and D. Gianola. 1992a. Bayesian inference about dispersion parameters of univariate mixed models with maternal effects: Theoretical considerations. *Genet. Sel. Evol.* 24:107–135.
- Cantet, R. J. C., L. R. Schaeffer, and C. Smith. 1992b. Reduced animal model with differential genetic grouping for direct and maternal effect. *J. Anim. Sci.* 70: 1730–1741.
- Cantet, R. J. C., D. Gianola, I. Misztal, and R. L. Fernando. 1993. Estimates of dispersion parameters and of genetic and environmental trends for weaning weight in Angus cattle using a maternal animal model with genetic grouping. *Livest. Prod. Sci.* 34:203–212.
- Cantet, R. J. C., A. N. Birchmeier, and J. P. Steibel. 2004. Full conjugate analysis of normal multiple traits with missing records using a generalized inverted Wishart distribution. *Genet. Sel. Evol.* 36:49–64.
- Cantet, R. J. C., A. N. Birchmeier, A. W. Canaza Cayo, and C. C. Fioretti. 2005. Semiparametric animal models via penalized splines as alternatives to models with contemporary groups. *J. Anim. Sci.* 83: 2482–2494.

- Cantet, R. J. C., and M. J. Suárez. 2009. Estimación bayesiana de matrices genéticas con distintos grados de credibilidad. 32° Congreso Argentino de Producción Animal. Rev. Arg. Prod. Anim. 29 (Supl. 1):6–8.
- Cappa, E. P., and R. J. C. Cantet. 2007. Bayesian estimation of a surface to account for a spatial trend using penalized splines in an individual-tree mixed model. Can. J. of Forest Res. 37:2677–2688.
- Carroll, R. J. 2005. Measurement error in epidemiologic studies. Encyclopedia of Biostatistics. New York, Wiley, 3:2491–2519.
- Carroll, R. J, D. Ruppert, and L. A. Stefanski. 1995. Measurement Error in Nonlinear Models. Chapman & Hall, London.
- Chib, S. 1993. Bayes regression with autoregressive errors: A Gibbs sampling approach. Journal of Econometrics. 58:275–294.
- Daniels, M. J., and M. Pourahmadi. 2002. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. Biometrika. 89:553–566.
- De Boor, C. 1972. On calculating with *B*-splines. J. Approximation Theory. 6:50–62.
- De Boor, C. 1993. *B*(asic)-spline basics. Fundamental Developments of Computer-Aided Geometric Modeling. L. Piegl, ed. Academic Press, San Diego, CA. Pages 27–49.
- Dickerson, G. E. 1947. Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. Research Bulletin. Iowa Agricultural Experimental Station 354:489–524.
- Dickey, D. A., and W. A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association. 74:427–431.
- Dickey, D. A., and W. A. Fuller. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica. 49:1057–1072.
- Dodenhoff, J., L. D. Van Vleck, S. D. Kachman, and R. M. Koch. 1998. Parameter estimates for direct, maternal and grandmaternal genetic effects for birth weight and weaning weight in Hereford cattle. J. Anim. Sci. 76:2521–2527.
- Dodenhoff, J., L. D. Van Vleck, and K. E Gregory. 1999a. Estimation of direct, maternal, and grandmaternal genetic effects for weaning weight in several breeds of beef cattle. J. Anim. Sci. 77:840–845.
- Dodenhoff, J., L. D. Van Vleck, and D. E. Wilson. 1999b. Comparison of models to estimate genetic effects for weaning weight of Angus cattle. J. Anim. Sci. 77:3176–3184.
- Durbán, M. 2009. An introduction to smoothing with penalties: *P*-splines. Boletín de Estadística e Investigación Operativa. 25:195–205.

- Eilers, P. H., and B. D. Marx. 1996. Flexible smoothing with *B*-splines and penalties. *Stat. Sci.* 11:89–121.
- Enders W. 1995. *Applied econometric time series*. J. Wiley, New York, USA.
- Foulley, J. L., and G. Lefort. 1978. Méthodes d'estimation des effets directs et maternels en sélection animale. *Ann. Génét. Sél. Anim.* 10:475–496.
- Fuller, W. A. 1987. *Measurement error models*. Wiley, New York.
- Gerstmayr, S. 1992. Impact of the data structure on the reliability of the estimated genetic parameters in an animal model with maternal effects. *J. Anim. Breed. Genet.* 109:321–336.
- Geyer, C. J. 1992. Practical Markov Chain Monte Carlo. *Stat. Sci.* 7:473–511.
- Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, L. M., J. O. Berger, A. P. Dawid, and A. F. M Smith editors. *Bayesian Statistics 4*. Oxford University Press. UK.
- Gutiérrez, J. P., I. Fernández, I. Alvarez, L. J. Royo, and F. Goyache. 2006. Sire x contemporary group interactions for birth weight and preweaning growth traits in the Asturiana de los Valles beef cattle breed. *Livest. Sci.* 99:61–68.
- Hamilton, J. D. 1994. *Time Series Analysis*. Princeton University Press, Princeton, New Jersey. USA.
- Heckman, N. 1986. Spline smoothing in partly linear models. *Journal of the Royal Statistical Society Ser. B.* 48:244–248.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447.
- Henderson, C. R. 1976. Rapid method for computing the inverse of relationship matrix. *Biometrics* 58: 1727–1730.
- Henderson, C. R. 1984. *Applications of linear models in animal breeding*. University of Guelph, Guelph, Ontario, Canada.
- Henderson, C. R. 1985. Equivalent linear models to reduce computations. *J. Dairy Sci.* 68:2267–2277.
- Henderson, C. R. 1988. Theoretical basis and computational methods for a number of different animal models. *J. Dairy Sci.* 71 (Suppl 2):1–16.
- Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2003. Genetic change for clinical mastitis in Norwegian cattle: a threshold model analysis. *J. Dairy Sci.* 86:369–375.
- Hobert, J. P., and G. Casella. 1996. The effects of improper priors on Gibbs sampling in hierarchical linear models. *J. Amer. Statist.* 91:1461–1473.

- Im, S., R. L. Fernando, and D. Gianola. 1989. Likelihood inferences in animal breeding under selection: a missing-data theory viewpoint. *Genetics Selection Evolution* 21:399–414.
- Iwaisaki, H., S. Tsuruta, I. Misztal, and J. K. Bertrand. 2005. Estimation of correlation between maternal permanent environmental effects of related dams in beef cattle. *J. Anim. Sci.* 83:537–542.
- Jensen, J., C. S. Wang, D. A. Sorensen, and D. Gianola. 1994. Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agric. Scand.* 44:193–201.
- Jiang J. 1996. REML Estimation asymptotic behavior and related topics. *Annals of Statistics* 24:225–286.
- Johnsson, I. D., and S. V. Morant. 1984. Evidence of a negative relationship between heifer growth and first calf weaning weight in commercial beef herds. *Aust. J. Exp. Agric. Anim. Husb.* 24:10–14.
- Koch, R. M. 1972. The role of maternal effects in animal breeding: VI. Maternal effects in beef cattle. *J. Anim. Sci.* 35:1316–1323.
- Kress, D. D., and P. J. Burfening. 1972. Weaning weight related to subsequent most probable producing ability in Hereford cows. *J. Anim. Sci.* 35:327–335.
- Lee, C., and E. J. Pollak. 1997. Relationship between sire \times year interactions and direct-maternal genetic correlation for weaning weight of Simmental cattle. *J. Anim. Sci.* 75:68–75.
- Ljung, G. M., and G. E. P. Box. 1978. On a measure of lack of fit in time series models. *Biometrika*. 65:297–303.
- Maniatis, N., and G. E. Pollott. 2003. The impact of data structure on genetic (co)variance components of early growth in sheep, estimated using an animal model with maternal effects. *J. Anim. Sci.* 81:101–108.
- Meyer, K. 1992. Variance Components due to direct and maternal effects for growth traits of Australian beef cattle. *Livest. Prod. Sci.* 31:179–203.
- Meyer, K. 1997. Estimates of genetic parameters for weaning weight of beef cattle accounting for direct-maternal environmental covariances. *Livest. Prod. Sci.* 52:187–199.
- Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). 7th World Congress on Genetics Applied to Livestock Production:19–23; Montpellier, France.
- Munilla Leguizamón, S., and R. J. C. Cantet. 2010. Estimation of residual dam-offspring correlation for a maternal animal model through a Griddy Gibbs Sampler. 9th World Congress on Genetics Applied to Livestock Production: 1–6. Leipzig, Germany.
- Munilla, S., and R. J. C. Cantet. 2012. Bayesian conjugate analysis using a generalized inverted Wishart distribution accounts for differential uncertainty among the genetic

- parameters- an application to the maternal animal model. *J. Anim. Breed. Genet.* 129: 173–187.
- O’Sullivan, F. 1986. A statistical perspective on ill-posed inverse problems (with discussion). *Statist. Sci.* 1:505–527.
- O’Sullivan, F. 1988. Fast computation of fully automated log-density and log-hazard estimators. *SIAM. J. Sci. Statist. Comput.* 9:363–379.
- Patterson, H. D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika.* 58:454–554.
- Pruzzo, L.; R. J. C. Cantet, and C. C. Fioretti. 2003. Risk-adjusted expected return for selection decisions. *J. Anim. Sci.* 814:2984–2988.
- Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32:949–953.
- Quaas, R. L. 1992. REML Note book. Mimeo, Dept Anim Sci, Cornell Univ., Ithaca, NY.
- Quaas, R. L., and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* 51:1277–1287.
- Quintanilla, R., L. Varona, M. R. Pujol, and J. Piedrafita. 1999. Maternal animal model with correlation between maternal environmental effects of related dams. *J. Anim. Sci.* 77:2904–2917.
- Reinsch, C. H. 1967. Smoothing by spline functions. *Numerische mathematik* 10:177–183.
- Rice, J. A, and C. O. Wu. 2001. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics.* 57:253–259.
- Robert, C. P. 1995. Simulation of truncated normal variables. *Statistics and Computing.* 5:121–125.
- Robert, C. P., and G. Casella. 1999. Monte Carlo Statistical Methods. Springer–Verlag. NY.
- Robinson, D. L. 1996. Models which might explain negative correlations between direct and maternal genetic effects. *Livest. Prod. Sci.* 45:111–122.
- Ruppert, D., and R. J. Carroll. 2000. Spatially-adaptative penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, 42 (2):205–224.
- Ruppert, D., M. P. Wand; and R. J. Carroll. 2003. Semiparametric regression. Cambridge Univ. Press, Cambridge, UK.
- SAS Institute Inc. 2008. SAS/ETS® 9.2. User’s guide. Cary, NC, USA.
- SAS Institute Inc. 2012. What’s new in SAS® 9.3. Cary, NC, USA.
- Schaeffer, L. R., and C. R. Henderson. 1983. Best linear unbiased prediction when error vector is correlated with other random vectors in the model. *Gen. Sel. Evol.* 15:395–400.
- Schaeffer, L. R., and B. W. Kennedy. 1989. Effects of embryo transfer in beef cattle on genetic evaluation methodology. *J. Anim. Sci.* 67:2536–2543.
- Searle, S. R. 1971. Linear models. J. Wiley & Sons, New York.

- Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52:591–611.
- Smith, B. J. 2003. Bayesian Output Analysis Program (BOA) version 1.0 user's manual. Disponible en <http://www.public-health.uiowa.edu/boa/Home.html>. Fecha de acceso: 20/10/2011.
- Smith, B. J. 2007. Boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software*, 21(11):1–37.
- Smith, C. 1984. Biotechnology in animal breeding programs. In: Second World Congress on Sheep and Beef Cattle Breeding. Paper 27. CSIR, Pretoria, S. Africa.
- Sorensen, D., and D. Gianola. 2002. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer–Verlag, NY.
- Speidel, S. E., R. M. Enns, and D. J. Garrick. 2007. Weaning weight inheritance in environments classified by maternal body weight change. *J. Animal Sci.* 85:610–617.
- Tanner, M. A. 1996. Tools for statistical inference, 3rd ed. Springer–Verlag, NY.
- Van Vleck, L. D. 1990a. Breeding value prediction with maternal genetic groups. *J. Anim. Sci.* 58:3998–4013.
- Van Vleck, L. D. 1990b. Alternative animal model with maternal effects and foster dams. *J. Anim. Sci.* 68:4026–4038.
- Van Vleck, L. D., E. J. Pollak, and E. A. B. Oltenacu. 1987. Genetics for the animal science. W. H. Freeman and Company. New York.
- Verbyla, A. P., B. R. Cullis, M. G. Kenward, and S. J. Welham. 1999. The analysis of designed experiments and longitudinal data using smoothing splines. *Appl. Stat.*, 48:269–312.
- Wade, K. M., and R. L. Quaas. 1993. Solutions to a system of equations involving a first-order autoregressive process. *J. Dairy Sci.* 76:3026–3032.
- Wade, K. M., R. L. Quaas, and L. D. Van Vleck. 1993. Estimation of the parameters in a first-order autoregressive process for contemporary groups. *J. Dairy Sci.* 76:3033–3040.
- Wand, M. P. 2003. Smoothing and mixed models. *Comput. Stat.* 18:223–249.
- Willham, R. L. 1963. The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* 19:18–27.
- Willham, R. L. 1972. The role of maternal effects in animal breeding: III. Biometrical aspects of maternal effects in animals. *J. Anim. Sci.* 35:1288–1293.
- White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82:632–638.
- Zellner, A. 1971. An introduction to Bayesian inference in econometrics. J. Wiley & Sons, New York, NY.